

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна
Факультет комп'ютерних наук
Кафедра теоретичної та прикладної системотехніки

«Затверджую»
Зав. кафедри теоретичної та
прикладної системотехніки
д.т.н., проф. С. І. Шматков
«___» _____ 2022 р

Пояснювальна записка

до кваліфікаційної роботи
магістра

на тему: «МОДЕЛІ ТРАФІКУ КОМП'ЮТЕРНИХ МЕРЕЖ»

Захищено на засіданні
Атестаційної комісії № 44
протокол № __ від __.12.2022 р.
Оцінка _____ / _____
Голова Атестаційної комісії
д.т.н., професор
_____ **МІНУХІН С.В.**

Виконав:

студент 2 курсу, групи КІ– 61
за спеціальністю 123 – Комп'ютерна
інженерія
Галузь знань: 12 – Інформаційні
технології


ЗАЦ Олександр Дмитрович 

Керівник:

д.т.н., професор, завідувач кафедри
теоретичної та прикладної
системотехніки

ШМАТКОВ Сергій Ігорович 

Рецензент:

к.ф.-м.н., доцент, в.о. зав. кафедри
електроніки і управляючих систем
ХРУСЛОВ Максим Михайлович 

Харків – 2022

АНОТАЦІЯ

Кваліфікаційна робота складається із вступу, трьох розділів, висновку, списку використаних джерел і 4 додатків. Загальний обсяг роботи складає 71 сторінки із яких 55 сторінок основної частини, 2 таблиці, 24 рисунки, 42 найменування списку використаних джерел на 5 сторінках і 4 додатків на 8 сторінках. Метою даної роботи є розробка найбільш перспективної моделі трафіку комп'ютерних мереж, яка б враховувала динамічні характеристики параметрів мереж.

У роботі були розглянуті найбільш розповсюджені моделі трафіку комп'ютерних мереж, їх особливості і обмеження у використанні. Також була створена програмна реалізація моделі самоподібного трафіку та проведений аналіз трафіку реальної комп'ютерної системи.

Програмна модель трафіку була протестована для різних параметрів комп'ютерної мережі, які обираються користувачем.

Створена програмна модель може бути використана у практичній діяльності адміністраторів та проектувальників комп'ютерних мереж для визначення параметрів мережевого обладнання.

Ключові слова: комп'ютерні мережі, трафік мереж, самоподібні моделі, нелінійні параметри мереж.

ABSTRACT

The thesis consists of an introduction, three chapters, a conclusion, a list of used sources and 4 appendices. The total volume of work is 71 pages, of which 55 pages of the main part, 2 tables, 24 figures, 42 names of the list of used sources on 5 pages and 4 appendices on 8 pages. The purpose of this work is to develop the most promising traffic model of computer networks, which would take into account the dynamic characteristics of network parameters.

The work considered the most common traffic models of computer networks, their features and limitations in their use. A software implementation of the self-similar traffic model was also created and an analysis of the real computer system traffic was carried out.

The software traffic model has been tested for various user-selectable computer network parameters.

The created software model can be used in the practical activities of computer networks administrators and designers to determine the network equipment parameters.

Keywords: computer networks, network traffic, self-similar models, nonlinear network parameters.

ЗМІСТ

ВСТУП	6
РОЗДІЛ 1. ОГЛЯД МОДЕЛЕЙ ТРАФІКУ КОМП'ЮТЕРНИХ МЕРЕЖ.....	8
1.1 Поняття мережевого трафіку	8
1.2 Класифікація моделей мережевого трафіку	9
1.2.1 Моделі, засновані на процесах відновлення	10
1.2.2 Моделі трафіку, засновані на марківських процесах	14
1.2.3 Авторегресійні моделі трафіку	17
1.2.4 Самоподібні моделі	24
1.2.5 Моделі трафіку додатків	28
Висновки за розділом 1	30
РОЗДІЛ 2. МОДЕЛІ САМОПОДІБНОГО ТРАФІКУ ТА МЕТОДИ ЇХ АНАЛІЗУ	
.....	32
2.1 Моделювання самоподібного трафіку	32
2.2 Показник самоподібності Херста і методи його оцінки	34
2.2.1 Метод R/S статистики.....	35
2.2.2 Метод флуктуаційного аналізу з виключенням тренду	36
2.3 Фрактальні параметри	38
2.4 Мережевий трафік як детермінований хаос	42
Висновки за розділом 2	44
РОЗДІЛ 3. МОДЕЛЮВАННЯ ТРАФІКУ КОМП'ЮТЕРНОЇ МЕРЕЖІ	45
3.1 Програмна реалізація моделі самоподібного трафіку	45
3.2 Аналіз реального трафіку мережі передачі даних	46
3.2.1 Опис набору даних.....	46
3.2.2 Статистичний аналіз параметрів трафіку	48

3.2.3 Аналіз нелінійних властивостей.....	55
Висновки за розділом 3	56
ВИСНОВКИ.....	57
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	59
ДОДАТКИ.....	64
Додаток А.....	64
Додаток Б	66
Додаток В.....	68
Додаток Г	71

ВСТУП

Швидкий і стрімкий розвиток мережевих і комунікаційних технологій, а також підвищення рівня доступності користування цими технологіями призводить до постійного зростання об'єму мережевого трафіку. Щорічно трафік у середньому зростає на 30-40%. При цьому більше 60% трафік складають мультимедійні застосунки потокового аудіо і відео. І саме такий трафік є більш чутливим до затримок у передачі пакетів даних. Таким чином, до мереж передачі даних постійно зростають вимоги до забезпечення якості зв'язку, рівню продуктивності і пропускної спроможності, якості управління трафіком, боротьби з перенавантаженням тощо.

Для того, щоб задовольнити вимоги до роботи мережі передачі даних необхідно визначити раціональний варіант її структури і параметрів. Пошук такого варіанта мережі є складною задачею, пов'язаною з вибором великої кількості різнорідних параметрів мережі – типів комутаційного обладнання, серверів, типів операційних систем та їх параметрів, стеків комутаційних протоколів та ін.

На сьогодні ефективного управління існуючою мережею і створення нової неможливо без детального попереднього моделювання її роботи.

Метою моделювання при цьому є визначення оптимальної топології, адекватний вибір мережевого обладнання, визначення робочих характеристик мережі тощо.

Існує два підходи до моделювання мереж та їх характеристик: імітаційний і аналітичний. За імітаційного моделювання використовують або готові спеціалізовані програмні системи, в яких процес створення моделі спрощений, або створюються програми за допомогою універсальних мов програмування. Системи імітаційного моделювання зазвичай містять набір засобів для

підготовки вихідних даних про досліджувану мережу (її топологію і трафік), засоби для статистичної обробки отриманих результатів моделювання. Якість результатів моделювання значно залежить від точності вихідних даних про мережу, які передаються до системи імітаційного моделювання.

Визначити ж характеристики мережі ще на стадії проектування можна за допомогою аналітичного моделювання. При побудові аналітичних моделей мережа і всі процеси, які в ній відбуваються, подаються у вигляді сукупності математичних співвідношень.

Об'єкт дослідження: процес моделювання характеристик комп'ютерних мереж.

Предмет дослідження: моделі характеристик комп'ютерних мереж і методи їх аналізу.

Мета дослідження: розробка найбільш перспективної моделі трафіку комп'ютерних мереж, яка б враховувала динамічні характеристики параметрів мереж.

Для досягнення сформульованої мети були виділені такі **задачі дослідження:**

- провести аналітичний огляд наявних моделей трафіку комп'ютерних мереж, визначити їх особливості використання;
- проаналізувати моделі самоподібного трафіку;
- ознайомитись з алгоритмами створення моделей самоподібного трафіку;
- створити програмну модель трафіку комп'ютерної мережі;
- виконати тестування моделі трафіку комп'ютерної мережі;
- провести аналіз реального трафіку комп'ютерної мережі та порівняти з модельованим.

РОЗДІЛ 1.

ОГЛЯД МОДЕЛЕЙ ТРАФІКУ КОМП'ЮТЕРНИХ МЕРЕЖ

1.1 Поняття мережевого трафіку

Слово «трафік» у перекладі з англійської (traffic) означає «рух», «перевезення», «транспорт», і в початковому сенсі мало значення об'єму дорожнього руху транспортних засобів крізь окремий вузол в одиницю часу.

Термін «трафік» з'явився у цифрових технологіях як поняття мобільного трафіку, тобто кількості секунд або хвилин стільникового зв'язку, яку використовує абонент оператора зв'язку.

Потім термін перейшов до сегменту мережевих технологій. Тут *трафік* – це об'єм інформації, який передається мережею за визначений час. У залежності від об'єкту, для якого трафік підраховується, це може бути інформація, яка пройшла як крізь окремий вузол (сервер), так і крізь цілі сегменти мережі.

За типом надходження до мережі розрізняють трафік:

- вхідний – який надходить до мережі, серверу, комп'ютеру;
- вихідний – який йде від мережі, серверу, комп'ютеру у зворотному напрямку;
- внутрішній – який являє собою рух інформації у середині визначеної мережі, наприклад, локальної або окремого сегменту глобальної, обмеженої за якоюсь ознакою;
- зовнішній – уся інформація за межами умовної мережі (локальної або окремого сегменту глобальної, обмеженої за якоюсь ознакою).

За часом спостереження трафік ділять на:

- добовий;

- місячний;
- річний.

За якістю роботи мережі трафік може бути нормальним або аномальним. Аномальний трафік – це трафік, не характерний для роботи мережі у звичайному режимі. Такий трафік може виникати, наприклад, при мережевих атаках.

В останній час у зв'язку зі збільшенням використання інтернету з'явилося поняття трафіку відвідування сайтів. Під ним розуміють кількість відвідувачів, які заходять на якийсь сайт за добу.

Мережевий трафік вимірюють у пакетах або в одиницях інформації: у бітах (кілобітах, мегабітах, гігабітах) чи у байтах (кілобайтах, мегабайтах, гігабайтах).

1.2 Класифікація моделей мережевого трафіку

У процесі моделювання мережі передачі даних важливо використовувати відповідну модель мережевого трафіку. Те, наскільки модель відповідає параметрам реальної мережі, визначає якість результатів моделювання.

У найпростішому випадку трафік можна подати як процес надходження дискретних даних (пакетів, повідомлень, одиничних сигналів тощо) і математично описати як послідовність даних, які надходять, $X_1, X_2, \dots, X_n, \dots$, де $X_0 = 0$. У такому випадку послідовністю даних може бути кількість даних, які надходять, і час між їхніми надходженнями [1].

Змішаний трафік буде містити більше одного елемента у послідовності даних X_n , яка надходить. Для опису змішаного трафіку використовується невід'ємна випадкова послідовність B_1, B_2, \dots, B_n , де $n = 1, 2, \dots, \infty$, B_n – кількість елементів у послідовності даних, які надходять.

Моделі трафіку можна класифікувати за характером процесу надходження даних і програмному забезпеченню або додатку, який здійснює передачу даних. За характером процесу моделі може бути стаціонарними й нестаціонарними.

Стационарні моделі, у свою чергу, можуть володіти короткостроковою або довгостроковою залежністю. До моделей із короткостроковою залежністю відносять класичні регресійні моделі і моделі, які основані на марківських процесах. Довгостроковою залежністю відрізняються фрактальні моделі [2]. Також за додатком-джерелом трафік може бути класифікований як трафік web, peer-to-peer, потокового відео тощо. На рис. 1.1 показана класифікація моделей мережевого трафіку [3].



Рисунок 1.1 – Класифікація моделей трафіку мереж передачі даних.

1.2.1 Моделі, засновані на процесах відновлення

Одними з перших розроблених моделей трафіку були моделі, засновані на теорії відновлення. Через свою простоту вони широко застосовувались у дослідженнях мереж передачі даних. Інтервали часу між подіями у процесі відновлення є додатними, незалежними і рівномірно розподіленими величинами. Процес відновлення можна визначити за допомогою процесу підрахунку

$\{N(t); t \geq 0\}$, де $N(t)$ – це кількість подій системи на інтервалі $(0; t]$. На кожному періоді настання подій $S_n = X_1 + \dots + X_n$ з визначеною ймовірністю процес починається знову. Тобто, якщо n -на подія настає при $S_n = \tau$, тоді, починаючи з $S_n = \tau$, j -та підпоследовність періоду настання подій: $S_{n+j} - S_n = X_{n+1} + \dots + X_{n+j}$. Таким чином, при $S_n = \tau$, $\{N(\tau + t) - N(\tau); t \geq 0\}$ – розраховується функція процесу відновлення з незалежними, рівномірно розподіленими інтервалами між подіями.

Процес відновлення нескладно використовувати, але він має істотний недолік – функція автокореляції ряду $\{X_n\}$ звертається в нуль для всіх ненульових лагів, що не відповідає результатом дослідження реального трафіку. Тобто аналіз автокореляційної функції (АКФ) у такому випадку говорить про відсутність часової залежності часового ряду. Більш того, додатне значення автокореляції ряду $\{X_n\}$ може пояснити наявність коротких спалахів мережевої активності [4]. Саме трафік змінного характеру, з періодами підвищеної активності, переважає у комп'ютерних мережах, тому модель, яка враховує автокорельованість даних, більш відповідає дійсній мережі.

Модель, заснована на розподіленні Пуассона. Модель трафіку на основі розподілення Пуассона є однією з перших і найбільш часто використовуваних. Вона застосовувалась для дослідження телефонних мереж.

Пуассоновський процес – це окремий випадок процесу відновлення, у якому настання подій експоненційно розподілене з параметром λ : $P\{X_n \leq t\} = 1 - e^{-\lambda t}$. Розподілення Пуассона варто застосовувати, коли трафік надходить від сукупності незалежних джерел, які відповідають вимогам розподілення. Середнє значення і дисперсія розподілення Пуассона визначаються параметром λ .

Графічно розподілення Пуассона може мати вигляд обмеженого біноміального розподілення. Розподілення володіє низкою математичних властивостей:

1) суперпозиція незалежних пуассоновських процесів дає новий пуассоновський процес з розподіленням, яке дорівнює сумі розподілень вихідних процесів;

2) властивість незалежних приростів усуває часові залежності ряду;

3) відповідно до теореми Пальма, пуассоновський процес частіше за все використовується для моделювання сукупності незалежних джерел трафіку.

Однак, агрегування трафіку не завжди призводить до розподілення Пуассона.

Функція розподілення ймовірності для пуассоновського процесу:

$$F(t) = 1 - e^{-\lambda t}. \quad (1.1)$$

Функція щільності розподілення:

$$f(t) = \lambda e^{-\lambda t}. \quad (1.2)$$

Найпростіший спосіб визначення того, що процес має пуассоновський характер – графічний. Для цього достатньо визначити, що гістограма часу настання подій зменшується за експоненціальним законом.

У випадку, якщо модель трафіку на основі розподілення Пуассона залежить від часу, тобто λ не є постійною величиною, то параметр розподілення є функцією від часу $\lambda(t)$.

Для моделювання глобальних мереж, у яких вага одного абонента щодо загальної картини трафіку мала, сесії користувачів можуть бути подані як пуассоновський процес. Також пуассоновський процес підходить до моделювання TCP трафіку на сесійному рівні моделі OSI, коли сесії ініціюються користувачем, тобто є TELNET і FTP додатки [5].

Модель на основі розподілення Бернуллі. Модель трафіку на основі розподілення Бернуллі – це дискретний аналог моделі Пуассона. Ймовірність p настання події у будь-який проміжок часу не залежить від інших подій. Для часового проміжку k відповідна кількість настання подій відповідає біноміальному розподіленню:

$$P\{N_k = n\} = \binom{k}{n} p^n (1 - p)^{k-n}, \quad (1.3)$$

де n приймає значення від 0 до k .

Час між настанням подій визначається параметром p з геометричним розподіленням

$$P\{A_n = j\} = p(1 - p)^j, \quad (1.4)$$

де j – додатне ціле число.

Модель на основі фазового процесу відновлення. Однією з моделей, заснованих на процесі відновлення, є модель трафіку фазового типу. Фазовий процес настання подій може бути змодельований як безперервний у часі марківський процес поглинання $C = \{C(t)\}_{t=0}^{\infty}$ у просторі допустимих станів $\{0, 1, \dots, m\}$, де 0 відповідає стану поглинання, а всі інші стани – перехідні, і при цьому процес поглинання виконується протягом кінцевого проміжку часу. Для визначення X_n процес C запускається з початковим розподіленням π . Коли відбувається поглинання, тобто процес входить у стан 0, процес зупиняється. Пройдений час буде відповідати X_n , що призведе до ймовірнісної комбінації сум експонент. Потім процес виконується знову з початковим розподіленням π , і процедура повторюється незалежно для отримання X_{n+1} .

Використання моделі на основі фазового процесу відновлення дозволяє керувати характеристиками мережі, яка моделюється, а також задовольняє умові апроксимації розподілень вхідних повідомлень [4].

1.2.2 Моделі трафіку, засновані на марківських процесах

Моделі трафіку, засновані на марківських процесах, вводять залежність між елементами у випадкову послідовність на відміну від моделей на основі процесів відновлення. Вважають, що ймовірність переходу системи у наступний стан S_{n+1} залежить лише від S_n і не залежить від інших станів S_i , де $i < n$. Це призводить до додатної автокореляції у $\{S_n\}$, що відповідає змінному характеру мережевого трафіку, коли періоди підвищеної активності йдуть за періодами зниженої інтенсивності передачі даних. У таких моделях кількість станів кінцева. Чим вона вища, тим більше модель відповідає реальній мережі передачі даних, однак це викликає підвищення складності моделювання.

Напівмарківські моделі отримують, коли час між виникненням станів підпорядковується випадковому розподіленню ймовірностей. Якщо час між змінами станів моделі не враховується зовсім, то процес вважається дискретним марківським ланцюгом.

ON-OFF і IPP-моделі. ON-OFF модель широко використовується для моделювання мереж передачі голосових даних [6]. Модель застосовується, коли необхідно врахувати скейлінговий характер мережевого трафіку. При цьому припускається лише два стани ON і OFF, а час переходу між станами розподілений за експоненційним законом [7]. Для мережі, в якій N статистично ідентичних і незалежних ON-OFF джерел, кожне джерело характеризується L – середньою кількістю пакетів, які передані за ON період, піковим значенням S і середнім r . Рівноважна ймовірність джерела у такому випадку може бути розрахована як $\gamma = r/S$.

Традиційна ON-OFF модель генерує процес, який є маже гаусівським процесом, що отриманий від значної кількості ON-OFF джерел.

У межах IPP-моделі трафіку (Interrupted Poisson Process) мережа може бути лише у двох станах. У ON стані мережа здійснює передачу даних у відповідності до розподілення Пуассона, у OFF стані передача даних не здійснюється (рис. 1.2).

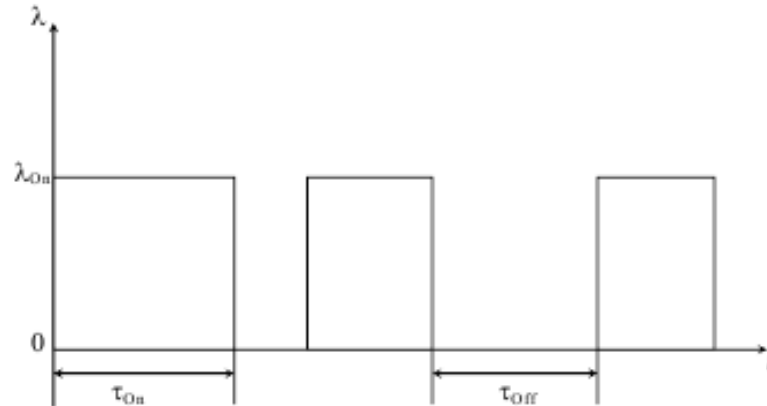


Рисунок 1.2 – Модель джерела трафіку.

Тривалість періодів ON і OFF є випадковими величинами. Довжина інтервалу активності τ_{ON} описується функцією щільності ймовірностей $p_{ON}(\tau)$, функцією розподілу

$$P_{ON}(\tau) = \int_0^{\tau} p_{ON}(t) dt. \quad (1.5)$$

Аналогічно задається довжина інтервалу пасивності. Зауважимо, що функція розподілу $P_{ON}(\tau)$ і $P_{OFF}(\tau)$ можуть бути різними для тривалості активного і пасивного періоду. При цьому швидкість потоку, який виробляється джерелом в активному стані дорівнює пропускній спроможності вхідного каналу зв'язку.

Модель на основі марківського процесу відновлення. У межах моделі на основі процесу відновлення для мережі є два стани: S_1 і S_2 . Амплітуда трафіку у стані S_1 дорівнює 0, у стані S_2 – 1. Якщо середні інтервали часу переходу між

станами прийняти рівними d_1 і d_2 відповідно, то ймовірність того, що система знаходиться у стані S_1 , дорівнює $P_{S_1} = \frac{d_1}{d_1+d_2}$, а у стані S_2 – відповідно $P_{S_2} = \frac{d_2}{d_1+d_2}$. При цьому суперпозиція незалежних процесів відновлення має біноміальне розподілення.

Модель на основі марківського модульованого пуассоновського процесу (ММПП). Через простоту реалізації пуассоновський процес сам по собі є достатньо привабливим способом моделювання мережевого трафіку. Однак, є очевидним недолік цього підходу через використання постійної швидкості потоку λ . Якщо розглянути дійсний трафік мовних даних, то швидкість потоку не є однаковою, оскільки повідомлення починаються і закінчуються у випадкові моменти часу.

Нехай N голосових повідомлень є мультиплексованим потоком, а кожне окреме повідомлення – незалежний пуассоновський процес. Таким чином, базовий процес – пуассоновський зі швидкістю $\lambda(t)$. Швидкість потоку моделюється як $\lambda(t) = n(t)\lambda$, де $n(t)$ – це кількість активних у даний момент часу передач голосових повідомлень. У такому випадку, $n(t)$ – стан безперервного у часі ланцюгу Маркова. ММПП зберігає деякі властивості відсутності часової залежності пуассоновського процесу і може бути проаналізована у межах марківської теорії.

Марківський модульований пуассоновський процес широко використовується при моделюванні трафіку завдяки високій гнучкості в якісному налаштування отриманої моделі, його також називають подвійним стохастичним процесом.

Простий приклад марківського модульованого пуассоновського процесу – це модель з двома станами: активним, з відповідним додатним параметром розподілення Пуассона, і вимкнутим станом, за якого параметр Пуассона дорівнює 0. Стан ON у такому випадку відповідає передачі звуку, а стан OFF –

тиші. Таку модель можна покращити агрегуванням множини незалежних джерел, кожний з яких характеризується ММПП з індивідуальним модулюючим марківським процесом.

Марківська модульована рідинна модель. Рідинні моделі визначають трафік як безперервний потік з параметром, який визначає швидкість цього потоку. Подібні моделі особливо підходять у випадках, коли вплив окремого пакету на мережу передачі даних незначний. Рідинні моделі відрізняються від традиційних точкових тим, що ігнорують дискретну природу пакетів даних [6]. Назва моделі походить від аналогії з впливом однієї молекули рідини у трубі з водою.

Обробка рідинних моделей достатньо проста і не потребує високих обчислювальних потужностей. Найчастіше використовується марківська рідинна модель, при цьому поточний стан марківського ланцюгу визначає швидкість потоку (трафіку). При моделюванні VBR відео використовується марківська модульована модель з постійною швидкістю, в якій стану S_k відповідає постійна швидкість λ_k . Вхідним змінним параметром мережі на основі рідинної моделі є швидкість потоку, на фоні якої можна проводити дослідження поведінки окремих вузлів мережі. Частіше за все рідинні моделі враховують ON-OFF природу джерел повідомлень, при цьому OFF період відповідає відсутності трафіку, а протягом ON періоду повідомлення надходять детерміновано з постійною швидкістю. Періоди не залежать один від одного, а їх розподілення підпорядковується експоненційному закону. Простота моделювання і аналітичного трактування робить модель достатньо популярною.

1.2.3 Авторегресійні моделі трафіку

В авторегресійних моделях значення випадкової величини X_n залежить від набору попередніх значень $\{X_k\}$, де $k < n$. Авторегресійна модель порядку p записується як $AR(p)$, а випадкова величина може бути визначена так:

$$X_k = r_1 X_{k-1} + r_2 X_{k-2} + \dots + r_p X_{k-p} + W_k, \quad (1.6)$$

де W_k – випадкова величина (білий шум), r_i – дійсні числа, X_t – корельовані випадкові величини.

Автокореляційна функція процесу $AR(p)$ являє собою загасаючу синусоїду. Дискретна авторегресійна модель порядку p генерує стаціонарну послідовність випадкових величин з розподіленням імовірності і АКФ як у авторегресійній моделі порядку p .

Одним із способів налагодження подібної моделі є вибір розміру вибірки для визначення нових випадкових значень змінних.

Авторегресійна лінійна модель. Авторегресійна модель широко застосовується при моделюванні VBR відео трафіку для розробки систем управління перенавантаження у високошвидкісних мережах для передачі мультимедіа даних. Подібна популярність виходить через характер відео даних, де на 1 секунду доводиться 30 кадрів і, як наслідок, відмінності у цих кадрах частіше всього незначні. Істотні відмінності між двома наступними кадрами відео ряду вносять відмінності сцен, які призведуть до росту об'єму переданих даних. Таким чином, відеоряд у межах однієї сцени, без різких стрибків у об'ємі трафіку, може бути змодельований за допомогою авторегресійних моделей, а для різких переходів між кадрами можна використати марківські ланцюги.

У роботі [8] відео трафік моделюється у рамках виразу:

$$X_n = Y_n + Z_n + V_n C_n, \quad (1.7)$$

де Y_n і Z_n – два незалежних $AR(1)$ процеси. Завдяки використанню одночасно двох процесів авторегресії вдається привести автокореляційну функцію отриманої моделі до вигляду, який відповідає реальному відео трафіку. Добуток

$V_n C_n$ – це стан марківського ланцюга і незалежної нормально розподіленої випадкової величини, який додається для обліку скачкоподібного росту навантаження при зміні сцен. Подібна модель застосовується для алгоритмів стиснення відео, в яких передаються лише зміни при переході між кадрами.

Дискретна авторегресійна модель. Дискретна авторегресійна модель порядку p генерує стаціонарну послідовність дискретних випадкових величин із випадковим розподіленням імовірності й АКФ як у авторегресійного процесу порядку p ($AR(p)$).

Дискретний авторегресійний процес першого порядку – це окремий випадок $DAR(p)$ процесу, який визначається на основі двох послідовностей незалежних випадкових величин $\{V_n\}$ і $\{Y_n\}$. Випадкова величина V_n приймає одно з двох значень, 0 або 1 з імовірністю $(1 - \rho)$ і ρ відповідно. Випадкова величина Y_n має дискретний набір станів, а в матриці переходів $P = \{Y_n = i\} = \pi(i)$. Випадкова величина X_n формується виходячи з виразу DAR -процесу першого порядку:

$$X_n = V_n X_{n-1} + (1 - V_n) Y_n. \quad (1.8)$$

DAR -процес першого порядку – ланцюг Маркова з дискретним набором станів і S матрицею переходів

$$P = \rho I + (1 - \rho) Q, \quad (1.9)$$

де I – це одинична матриця, Q – матриця з $Q_{ij} = \pi(j)$ для $i, j \in S$.

DAR -процес першого порядку володіє АКФ авторегресійного процесу першого порядку, функція розподілення ймовірності – це функція від π .

$DAR(1)$ процес володіє меншою у порівнянні з ланцюгом Маркова, кількістю параметрів, а оцінка цих параметрів не викликає труднощів. Модель трафіку на основі DAR -процесу легко піддається аналітичному аналізу, однак через

експоненційне згасання функції автокореляції модель не можна застосувати до трафіку з повільно згасаючою АКФ.

Авторегресійна модель ковзного середнього. Авторегресійна модель ковзного середнього порядку (p, q) позначається як $ARMA(p, q)$ і має такий вигляд

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p}.$$

Що є еквівалентним запису

$$\varphi(B)X_t = c + \theta(B)\varepsilon_t, \quad (1.10)$$

де B – оператор лагу (оператор зсуву), такий що $X_{t-1} = BX_t$, а $\varphi(B) = 1 - \varphi_1 B^1 - \dots - \varphi_p B^p$, c – константа.

Це є еквівалентним фільтрації білого шуму ε_t лінійним фільтром, інваріантним до зсуву за часом, який має дрібно-раціональну передавальну функцію з p полюсами і q нулями [12], тобто

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{1 - \sum_{k=0}^q \theta_k z^{-k}}{1 - \sum_{k=1}^p \varphi_k z^{-k}}. \quad (1.11)$$

Автоковаріація $ARMA(p, q)$ процесу може бути отримана добутком $H(z)$ і X_{t-k} з урахуванням математичного сподівання і взаємної кореляції між ε_t і X_t .

$ARMA$ моделі широко застосовуються для моделювання VBR трафіку. У цьому випадку довжина відеокадру ділиться на рівні m інтервалів. Кількість комірок $n_i, i = 0, \dots, m - 1$, у часовому інтервалі моделюється за допомогою $ARMA$ процесу:

$$X_n = \varphi X_{n-m} + \sum_{i=0}^{m-1} \theta_i \varepsilon_{n-i}. \quad (1.12)$$

Оскільки відеодані кожного кадру корельовані між собою зі змінним коефіцієнтом кореляції, то функція автокореляції буде містити піки на лагах, кратних m . У розглядуваній моделі AR -частина використовується для ефекту повторної кореляції, а θ_k вводиться для підбору кореляції для інших затримок [12]. Параметричне оцінювання $ARMA$ моделей складніше, ніж AR моделей, оцінка θ_k потребує розв'язання множини нелінійних рівнянь. Аналітичний аналіз також є важким.

Інтегральна модель авторегресії ковзного середнього. Авторегресійний інтегральний процес ковзного середнього порядку (p, d, q) позначається як $ARIMA(p, d, q)$ і будується на основі $ARMA(p, q)$. Тобто $ARIMA(p, d, q)$ можна інтерпретувати як $ARMA(p+d, q)$ – модель з d одиничними коренями, інші корені багаточлену знаходяться поза межами одиничного кола. $ARIMA(p, d, q)$ процес можна записати так:

$$\varphi(B)\nabla^d X_t = \theta(B)\varepsilon_t, \quad (1.13)$$

де ∇ – оператор диференціювання, такий що $\nabla X_t = X_t - X_{t-1}$ і $\nabla X_t = (1 - B)X_t$, а $\varphi(B)$ – багаточлен від B .

$ARIMA(p, d, q)$ процес використовується для моделювання нестационарних рядів, які виявляють однорідність на відміну від їх локального рівня або тренду. Як правило, використовують $ARIMA$ з d , що дорівнює 1 або 0: якщо $d=0$, то процес має лінійний тренд, а якщо $d>1$, то поліноміальний.

Модель розширення і перетворення вибірки. Модель розширення і перетворення вибірки (TES) – це нелінійна регресійна модель, яка спрямована на відтворення наданої стаціонарної вибірки з дотриманням маргінального розподілення і структури АКФ [9]. Іншими словами, TES генерує послідовність, яка керується випадковим процесом, згладжену і адаптовану під дане маргінальне розподілення. Якість моделювання трафіку телекомунікаційних додатків залежить від того, наскільки змодельований часовий ряд відповідає

основним характеристикам вхідної послідовності. Найбільш важливих характеристик три: маргінальне розподілення, структура АКФ, відповідність між графічним поданням вихідних даних і змодельованого ряду. Загальний алгоритм моделювання показаний на рис. 1.3.

На першому етапі відбувається ініціалізація ряду $\{Z_i\} \in [0,1]$. Після чого Z_i використовується для вибору $V_i \in [-0.5,0.5]$ у відповідності до заданої щільності. Величина V_i додається до значення фонові послідовності Y_{i-1} і формується таким чином $Y_i = |Y_{i-1} + V_i|_{mod=1}$. У межах алгоритму перетворення фонові послідовності відбувається двічі. По-перше, розрахунки за модулем 1 Y_i призводять до деякого ефекту розриву, якщо припустити, що значення Y_i можуть бути достатньо близькими до 1. Також невеликі значення V_{i+1} можуть привести до близькості Y_{i+1} до нуля, тобто вагомим змінам у фоновій послідовності. Подібні ефекти усуваються функцією об'єднання (віконною функцією) S , яка значення близькі до 0 і 1 встановлює біля 0 (рис.1.4).

Друге перетворення полягає в інверсії гістограми емпіричних даних, що приводить об'єднану послідовність ближче до діапазону значень вихідного ряду і встановлює необхідне маргінальне розподілення. Це останній етап моделювання емпіричного часового ряду, який генерую основну послідовність. Однією з основних особливостей TES моделі є можливість подальших розрахунків лагу АКФ спотворення і щільності розподілення нововведень.

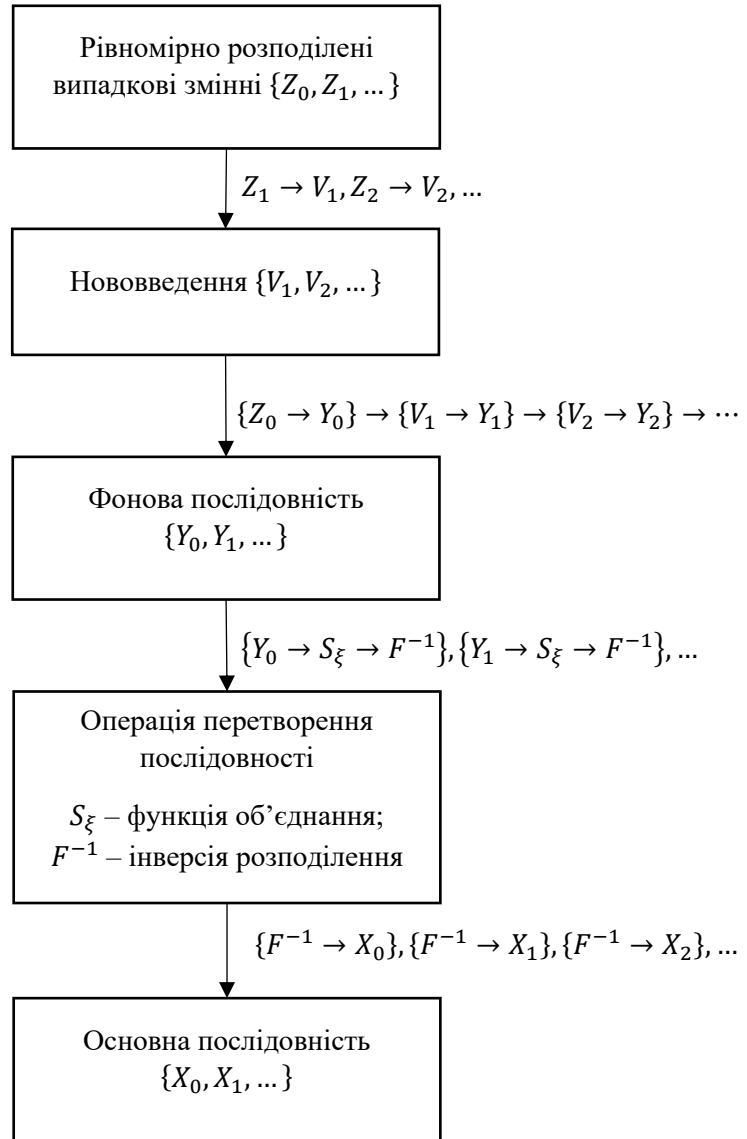


Рисунок 1.3 – Алгоритм TES.

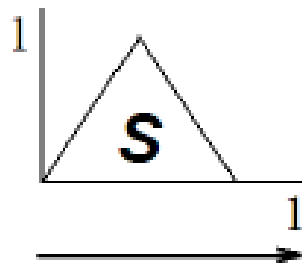


Рисунок 1.4 – Функція об'єднання.

1.2.4 Самоподібні моделі

Нехай $\{X_t\}$ – стаціонарний у широкому сенсі стохастичний процес, якій має стаціонарне середнє $\mu = E[X_t]$, стаціонарну і кінцеву дисперсію $\nu = E[(X_t - \mu)^2]$ і стаціонарну АКФ $\gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)]$, яка залежить лише від k і не залежить від t . Відмітимо, що $\nu = \gamma_0$. Якщо прийняти $\{X_k\}$ при лагу k як ρ_k , то за визначенням $\rho_k = \gamma_k/\gamma_0$.

Поділимо вихідний часовий ряд $\{X_t\}$ на послідовності, які не перетинаються, довжини m і усереднимо їх, щоб отримати послідовність $\{X_j^{(m)}\}$:

$$X_j^{(m)} = m^{-1}(X_{jm-m+1} + \dots + X_{jm}), \quad (1.14)$$

де $X_j^{(m)}$ – це вибіркове середнє послідовності $X_{jm-m+1} + \dots + X_{jm}$.

Позначимо ν_m як дисперсію $\{X_j^{(m)}\}$:

$$\begin{aligned} \nu_m &= E \left[\frac{1}{m} (X_{jm-m+1} + \dots + X_{jm}) \right]^2 \left[E \frac{1}{m} (X_{jm-m+1} + \dots + X_{jm}) \right]^2 = \frac{\nu}{m} + \\ &\quad \frac{2}{m^2} \sum_{k=1}^m (m-k) \gamma_k = \nu \left[1 + 2 \sum_{k=1}^m \left(1 - \frac{k}{m} \right) \rho_k \right] m^{-1}. \end{aligned}$$

У такому випадку, якщо процес є білим шумом, то $X_{jm-m+1} + \dots + X_{jm}$ будуть взаємно некорельованими, $\rho_k = 0$ для $k > 0$ і $\nu_m = \nu m^{-1}$.

Для великих m можна усереднити:

$$\nu_m = \nu [2 \sum_{k=1}^m \rho_k] m^{-1}. \quad (1.15)$$

Розглянемо випадок, коли $\rho_k \neq 0$ і $\sum_{k=-\infty}^{\infty} \rho_k < 0$. Тоді дисперсія вибіркового середнього буде асимптотично згасати і наближатися до нуля пропорційно m^{-1} , тобто

$$v_m \approx v c_p m^{-1}, \quad (1.16)$$

де $c_p = \text{const}$.

Для більшості розглянутих моделей, таких як ARMA або моделей на основі марківських процесів, середня дисперсія вибірки згасає саме у відповідності до останньої формули v_m .

У роботі [10] для дослідження трафіку вказане згасання середньої дисперсії вибірки навіть повільніше за m^{-1} . Найбільш простим підходом у даному випадку було б прийняти згасання v_m пропорційним до $m^{-\alpha}$ для деякої $\alpha \in (0,1)$. Тоді $\sum_{k=1}^m \rho_k$ повинна бути пропорційною до $m^{1-\alpha}$:

$$\sum_{k=1}^m \rho_k \approx C m^{1-\alpha}. \quad (1.17)$$

Оскільки $\alpha < 1$, то $\sum_{k=1}^{\infty} \rho_k \rightarrow \infty$. Таким чином АКФ згасає повільніше, тому що вона не додається [8].

Процес $\{X_t\}$ має короткострокову залежність, якщо $\sum_k \rho_k < \infty$. Відповідно, v_m згасає асимптотично пропорційно до $m^{-\alpha}$, спектральна щільність потужності у нулі звертається у нескінченність, а усереднені послідовності $\{X_t^{(m)}\}$ наближаються до чистого шуму при $m \rightarrow \infty$. АКФ процесів, які мають короткострокову залежність, згасає експоненційно.

Процес $\{X_t\}$ має довгострокову залежність, якщо $\sum_k \rho_k \rightarrow \infty$. Дисперсія середнього v_m згасає повільніше ніж m^{-1} . При цьому довгострокова залежність лише визначає поведінку АКФ при великих лагах, але не визначає автокореляцію ряду для будь-якого фіксованого кінцевого лагу.

Процес $\{X_t\}$ є самоподібним, якщо $\rho_k^{(m)} = \rho_k$ для усіх m і k , тобто структура АКФ зберігається на різних часових масштабах. Стохастичний самоподібний процес зберігає однакові статистичні показники на будь-якому діапазоні масштабів і задовольняє такому виразу:

$$\{X_{\alpha t}\}^D = \alpha^H \{X_t\}, \quad (1.18)$$

де D означає рівність у розподіленні, а H називається параметром Херста [15].

Якщо процес $\{X_t\}$ має стаціонарний інкремент $Y_t = X_t - X_{t-1}$, то АКФ:

$$\rho_k \rightarrow H(2H - 1)k^{2H-2} \text{ при } k \rightarrow \infty, \quad (1.19)$$

де $H = 1 - \frac{\alpha}{2}$, а при $0 < H < 1$ і $H \neq 0.5$ $\sum_k \rho_k \rightarrow \infty$.

Фрактальна інтегрована модель авторегресії – ковзного середнього. Фрактальна інтегрована модель авторегресії – ковзного середнього – *ARIFMA(p,d,q)* при $0 < d < 0.5$ – це приклад моделі стаціонарного процесу з довгостроковою залежністю. Вона є розширенням *ARIMA(p,d,q)* моделі і визначається так:

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t, \quad (1.20)$$

де $0 < d < 0.5$. Оператор $\nabla^d = (1 - B)^d$ можна виразити через біноміальне розподілення

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k, \quad (1.21)$$

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}, \quad (1.22)$$

де $\Gamma(x)$ – гамма-функція.

Оскільки гамма-функція має полюси для від'ємних цілих чисел, і біноміальні коефіцієнти дорівнюють нулю, якщо $k > d$ і d – ціле число, то для всіх додатних цілих чисел, лише перші $d+1$ будуть ненульовими.

За допомогою *ARIFMA* можна моделювати процеси з короткостроковою і довгостроковою залежністю.

Модель фрактального броунівського руху. Броунівський рух – це стохастичний процес $\{B_t\}$ для $t \geq 0$. Має такі властивості:

- інкремент $(B_{t+t_0} - B_{t_0})$ має нормальне розподілення з нульовим середнім і дисперсією $\sigma^2 t^{2H}$;
- інкремент на двох інтервалах, які не перетинаються, $[t_1, t_2]$ і $[t_3, t_4]$, $B_{t_4} - B_{t_3}$ і $B_{t_2} - B_{t_1}$ – незалежні випадкові величини;
- $B_0 = 0$ і B_t безперервна при $t = 0$.

Фрактальний броунівський рух $\{fB_t\}$ – це самоподібний гаусівський процес з параметром самоподібності $0.5 \leq H < 1$. Фрактальний броунівський рух відрізняється від броунівського руху тим, що дисперсія інкременту дорівнює $\sigma^2 t^{2H}$. Визначимо дисперсію інкременту

$$\sigma^2 = E\{(fB_t - fB_{t-1})^2\} = E\{(fB_1 - fB_0)^2\} = E\{fB_1^2\},$$

тоді

$$E\{(fB_{t_2} - fB_{t_1})^2\} = E\{(fB_{t_2-t_1} - fB_0)^2\} = \sigma^2 (t_2 - t_1)^{2H}.$$

Також

$$\begin{aligned} E\{(fB_{t_2} - fB_{t_1})^2\} &= E\{fB_{t_2}^2\} + E\{fB_{t_1}^2\} - 2E\{fB_{t_2}fB_{t_1}\} \\ &= \sigma^2 t_2^{2H} + \sigma^2 t_1^{2H} - 2\gamma(fB_{t_1}, fB_{t_2}), \\ \gamma(fB_{t_1}, fB_{t_2}) &= \frac{1}{2}\sigma^2 (t_2^{2H} - (t_2 - t_1)^{2H} + t_1^{2H}). \end{aligned}$$

Тому коваріація інкременту на двох інтервалах, які не перетинаються:

$$\begin{aligned} \gamma(fB_{t_4} - fB_{t_3}, fB_{t_2} - fB_{t_1}) &= \gamma(fB_{t_4}, fB_{t_2}) - \gamma(fB_{t_4}, fB_{t_1}) - \gamma(fB_{t_3}, fB_{t_2}) + \\ &\gamma(fB_{t_3}, fB_{t_1}) = \frac{\sigma^2}{2} (t_4 - t_1)^{2H} - (t_3 - t_1)^{2H} + (t_3 - t_2)^{2H} - (t_4 - t_2)^{2H}. \end{aligned}$$

Фрактальний броунівський рух $\{fB_t\}$ можна отримати з броунівського руху $\{B_t\}$ інтегруванням

$$fB_t = \int_0^t (t-u)^{H-0.5} dB(u), \quad (1.23)$$

таким чином, взаємозалежність між інкрементами фрактального броунівського руху може бути нескінченною.

У випадку дискретного процесу автокореляції інкременту може бути розрахована заміною t_1, t_2, t_3, t_4 на $n, n+1, n+k, n+k+1$ і діленням на σ^2 :

$$\rho_k = \frac{1}{2} [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]. \quad (1.24)$$

Процес інкрементування у даному випадку – це фрактальний гаусівський шум. Автокореляція в останньому виразі має довгострокову залежність, $\rho_k \sim k^{2H-2}, k \rightarrow \infty$. Один з прикладів моделювання трафіку за допомогою фрактального гаусівського шуму наведений у роботі [14], де також вказується на складність аналітичного аналізу розподілення заповнення буферу. Тому наведений лише наближений аналіз поведінки хвостів розподілень. Показано, що для великих значень H ріст навантаження на мережу потребує значного збільшення об'ємів пам'яті. Ймовірність втрати пакету зростає алгебраїчно з розміром буферу, а не експоненційно, як у випадку з марківськими і ARMA моделями.

1.2.5 Моделі трафіку додатків

Процес передачі даних у мережі та характер самих даних визначається додатками, які використовуються. Більшість додатків можна віднести до таких типів: Web, e-mail, peer-to-peer файлообмінники, потокове мультимедіа. У роботі

[12] говориться, що більше 40% усіх даних, які передаються мережею, відносяться до Web.

Можна застосувати розглянуті підходи для моделювання трафіку додатків, однак, краще використовувати спеціальні моделі, щоб отримати кращі результати.

Web-трафік. У деяких роботах [16, 17] згадується, що Web-трафік займає більшу частину всього трафіку мережі Internet, що не дивує з урахуванням того, що Web-браузери надають дружелюбний інтерфейс для роботи з електронною поштою, передачі файлів, віддаленої обробки даних, потокового медіа тощо.

Деякі дослідження Web-трафіку на самоподібність вказують коефіцієнт Херста близький до 0.8 [10]. Крім характеру самого трафіку було вказано, що Web-клієнти можуть бути подані як ON-OFF джерела повідомлень, які мають розподілення з важким хвостом. При цьому самоподібність трафіку виникає при агрегуванні множини потоків даних з розподіленням з важким хвостом. Щільність розподілення ймовірності, яка відповідає RTX, повинна мати степеневий характер. «Важкість» хвосту означає, що надвеликі значення ряду мають ненульову ймовірність на гістограмі розподілення. OFF періоди можуть бути або активними, коли завантаження великих об'ємів даних не відбувається, але користувач працює з Web-браузером, або неактивним, коли користувач не використовує Web-браузер. Активні OFF періоди можуть бути описані розподіленням Вейбулла, а неактивні – Парето розподіленням з важким хвостом.

У роботі [18] проводилось дослідження 1900 клієнтів Web-браузерів, для яких також була характерною ON-OFF модель. OFF період описувався розподіленням Вейбулла і йому відповідав період, за час якого користувач переглядає завантажений контент. ON період відповідає завантаженню сторінки, однак процес розглядався більш детально. Завантаження основного коду сторінки і додаткових об'єктів описувались окремими гамма-розподіленнями.

У роботі [19] проводився добовий збір трафіку з роутерів корпоративної мережі. За результатами дослідження, 42% всього потоку даних належало Web-трафіку. Були розглянуті щільності розподілення кількості серверів, з якими обмінювався даними кожний клієнт, а також кількість клієнтів, які обслуговувались одним сервером, об'єми трафіку, передані кожним клієнтом, об'єми переданих серверами даних. Одним з висновків став факт, що дуже складно виокремити якусь типову поведінку окремого клієнту або серверу.

Трафік однорангових мереж. Peer-to-peer (p2p) або трафік однорангових мереж часто порівнюють з Web-трафіком, оскільки однорангові мережі являють собою повну протилежність клієнт-серверній архітектурі.

У роботі [20] проводилось дослідження університетської мережі. Більша частина даних, які передавались, відносилась до відео-форматів avi, mpeg і mp3. При цьому найбільше навантаження давали декілька комп'ютерів у мережі, на яких були розташовані файли більші 700 Мб, та доступ до яких здійснювався іншими членами мережі. Таким чином, невелика кількість вузлів у мережі є відповідальною за основну частину навантаження.

У роботі [21] здійснювався збір p2p трафіку через роутери корпоративної мережі. І так само було встановлено, що найбільше навантаження у мережі створювали декілька комп'ютерів.

Висновки за розділом 1

Розвиток інформаційних технологій надав поштовх для розробки великої кількості моделей трафіку мереж передачі даних. Ранні моделі трафіку, такі як моделі Пуассона й їх модифікації, марківські моделі, припускали, що трафік мереж має тенденцію до згладжування сплесків, що виникають на окремих інтервалах часу, із зростанням кількості джерел трафіку, а також, що джерела трафіку мають незалежний характер. Але у сучасних умовах спільної активності

користувачів, які інтенсивно обмінюються інформацією між собою, використання таких моделей не дає необхідних результатів.

Обмін інформацією проявляється у формі нерегулярного характеру флуктуацій трафіку, що вказує на високі втрати і затримки пакетів у мережах. Тому були розроблені моделі, які описують нерегулярну динаміку мереж передачі даних. До них відносять ON/OFF моделі, фрактальний броунівський рух, фрактальний гаусівський шум, регресивні моделі трафіку (моделі на основі узагальнень лінійних авторегресійних процесів – AR, процеси змінного середнього – MA, авторегресійні моделі змінного середнього – ARMA, фрактальний авторегресійний інтегральний процес змінного середнього – FARIMA), фрактальні точкові процеси (Fractal point process) та інші. Кожна з цих моделей має власні переваги і недоліки. Наприклад, модель FARIMA дозволяє управляти кореляційною структурою і охоплювати як короткочасну, так і довготривалу залежність, але є більш складною в обчислюваннях, і вимагає більшого обсягу вихідних статистичних даних для її параметризації.

У роботі пропонується розглянути моделі трафіку мереж, які б враховували властивості самоподібності.

РОЗДІЛ 2. МОДЕЛІ САМОПОДІБНОГО ТРАФІКУ ТА МЕТОДИ ЇХ АНАЛІЗУ

2.1 Моделювання самоподібного трафіку

Вперше про самоподібність трафіку почали говорити на початку 90х років ХХ ст. у зв'язку зі збільшенням об'єму досліджуваних даних. У статті [28] вказується, що об'єднаний з декількох джерел трафік стає сильно автокорельованим з довгостроковою залежністю. Така поведінка може бути пояснена тим, що майбутнє процесу визначається його минулою поведінкою, при чому степінь впливу минулого на процес зменшується у часі. Сукупність множини джерел даних, які виявляють нескінченну дисперсію, в результаті дає самоподібний об'єднаний мережевий трафік.

Таким чином більш доречним для моделювання мережевого трафіку є використання самоподібних моделей. Перший тип моделей трафіку – це *фрактальні моделі* процесів, які оперують потоками пакетів і безперервним часом.

Фрактальний точковий процес або FSNDP (Fractal Shot Noise Driven Poisson process) процес відноситься до класу подвійних стохастичних процесів. Змінна у часі інтенсивність цього процесу задається безперервним випадковим процесом – фрактальним дробовим шумом, який отримується шляхом фільтрування класичного пуассоновського процесу [27].

Процес FSNDP повністю визначається набором параметрів $\langle \mu, \beta, K, A, B \rangle$.

Первісний пуассоновський потік $\{t_i\}$ з постійною інтенсивністю μ є входом для лінійного фільтру з імпульсною функцією

$$h(t) = \begin{cases} Kt^{-\beta}, & t \in (A, B), \\ 0, & t \notin (A, B), \end{cases} \quad (2.1)$$

де β визначається степенем самоподібності процесу, A, B – невід’ємні обмежувальні параметри, K – додатна константа, яка визначає амплітуду результуючого процесу. Фільтр викликає фрактальний дробовий шум:

$$I(t) = \sum_{i=-\infty}^{\infty} h(t - t_i), \quad (2.2)$$

який розглядається як змінна інтенсивність для другого пуассоновського точкового процесу, виходом якого є потік FSNDP (рис. 2.1).

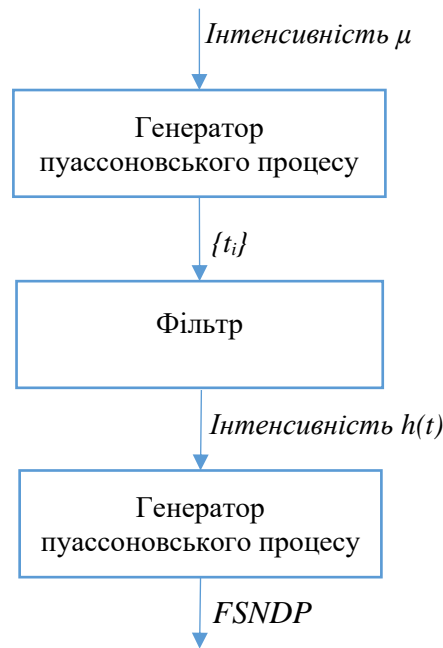


Рисунок 2.1 – Процес отримання FSNDP потоку.

Параметр β визначається степенем самоподібності процесу і пов’язаний з параметром Херста H :

$$H = \frac{3}{2} - \beta. \quad (2.3)$$

Процес має стійкі самоподібні властивості за умови $A \ll B$ і A , достатньо близьким до нуля. На практиці зазвичай приймають $A = 0 \dots 0,1$ і $B < 10$.

Другий тип моделей засновується на використанні фрактального гаусівського шуму або фрактального броунівського руху. Фрактальний броунівський рух $\{fB_t\}$ – це самоподібний гаусівський процес з параметром самоподібності $0.5 \leq H < 1$.

2.2 Показник самоподібності Херста і методи його оцінки

Коефіцієнт або показник Херста є найважливішим параметром, який характеризує степінь самоподібності [26, 29]. Оцінка параметру Херста допомагає не лише вирішити чи є процес самоподібним, але й дозволяє застосувати до процесу ряд методів з прогнозування фрактальних процесів.

Коефіцієнт Херста H приймає значення від 0 до 1:

- при значенні коефіцієнту $0.5 < H < 1$ говорять про персистентний (яка підтримується) поведінці процесу, або про те, що процес має довготривалу пам'ять, тобто є самоподібним. Персистентні стохастичні процеси виявляють чітко виражені тенденції змін при відносно малому шумі;

- коли $H = 0.5$ говорять про повністю випадковий ряд, який є аналогом зміщенню частинки при класичному броунівському русі;

- коли $0 < H < 0.5$ говорять про антиперсистентний процес. Такий ряд не має самоподібності.

Існують різні методи оцінювання параметру H для того, щоб визначити самоподібність або повільно спадаючу залежність.

Часові методи оцінки параметру Херста:

- метод R/S статистики, в основі якого лежить формула А. Ейнштейна про броунівських рух часток;

- метод варіацій;
- метод абсолютного моменту;
- метод відношення варіації залишків;
- частотні методи оцінки параметру Херста;
- флуктуаційний аналіз;
- періодограмний аналіз;
- оцінка на базі Вейвлет.

2.2.1 Метод R/S статистики

У дослідженнях трафіку часто використовують *метод R/S статистики*. Розглянемо його детальніше.

Показник Херста H визначається у термінах асимптотичної поведінки масштабованого діапазону як функція відрізка часу часового ряду таким чином:

$$E \left[\frac{R(n)}{S(n)} \right] = C \cdot n^H, n \rightarrow \infty, \quad (2.4)$$

де R – розмах накопичених відхилень n значень від середнього значення ряду, S – середнє квадратичне відхилення, n – величина проміжку часу (кількість точок на відрізка часового ряду), C – константа.

Даний метод легко застосовується на практиці, має достатній рівень надійності, але може бути чутливим до довжини часового ряду. Чутливість полягає у тому, що метод погано оцінює локальний показник Херста.

Алгоритм R/S аналізу:

1. Заданий вихідний часовий ряд Y_t . Розраховується логарифмічне відношення $L_t = \ln \frac{Y_t}{Y_{t-1}}$.

2. В отриманому ряді L_t послідовно виділяють K відрізків довжини n . Позначають отримані відрізки $L_k, k = 1 \dots K$.

3. Для кожного з цих відрізків обчислюється середнє значення

$$E[L_k] = \frac{1}{n} \sum_{i=1}^n L_{k,i}, k = 1 \dots K. \quad (2.5)$$

4. Розраховують розмах накопичених відхилень для кожного з відрізків

$$X_{k,t} = \sum_{i=1}^t (L_{k,i} - E[L_k]), t = 1 \dots n, k = 1 \dots K. \quad (2.6)$$

$$R_k = \max_{t=1 \dots n} (X_{k,t}) - \min_{t=1 \dots n} (X_{k,t}), k = 1 \dots K. \quad (2.7)$$

5. Обчислюють середнє квадратичне відхилення для кожного відрізка

$$S_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (L_{k,i} - E[L_k])^2}, k = 1 \dots K. \quad (2.8)$$

6. Обчислюють значення статистики

$$\frac{R}{S}(n) = \frac{1}{K} \sum_{k=0}^{K-1} \frac{R_k}{S_k}. \quad (2.9)$$

7. Повторюють алгоритм для наступного значення n .

Таким чином отримують послідовність точок

$$H(n) = \frac{\ln\left(\frac{R(n)}{S(n)}\right)}{\ln(n)}, \quad (2.10)$$

де $n = N, \frac{N}{2}, \frac{N}{4}, \dots, N$ – кількість елементів у вихідному часовому ряді.

Далі розв'язують задачу регресії $\ln\left(\frac{R}{S}(n)\right) = H \log(n) + c$ за відомими точками $(H(n), n)$, будують лінію тренду та знаходять апроксимоване значення показника Херста [26].

2.2.2 Метод флуктуаційного аналізу з виключенням тренду

Метод R/S статистики має ряд суттєвих обмежень, особливо при спостереженні трафіку на фоні шумів, трендів, періодичних складових. Для запобігання цим обмеженням використовують *метод флуктуаційного аналізу з виключенням тренду* (Detrended Fluctuation Analysis, DFA) [31].

Нехай є ряд даних $x_i, i = 1, \dots, N$. Наприклад, для добового запису трафіку мережі значення x_i можуть бути послідовністю числа отриманих пакетів або байт в кожному короткотривалому фрагменті T (секунда, хвилина), у випадку $T=1\text{с}$, довжина послідовності $N=24 \times 60 \times 60=86400$.

Алгоритм аналізу ряду даних з використанням DFA складається з чотирьох кроків.

1. Визначення кумулятивного ряду

$$Y(i) = \sum_{k=1}^i x_k - \langle x \rangle, \quad (2.11)$$

де $\langle x \rangle$ є середнім значенням: $\langle x \rangle = \frac{1}{N \sum_{i=1}^N x_i}$.

2. Розбиття ряду $Y(i)$ на $N_s = \lceil N/s \rceil$ непересічних сегментів тривалістю s . Оскільки може виявитися, що довжина ряду N не буде кратною довжині вікна s , то у такому випадку залишається коротка частина в кінці ряду $Y(i)$. Для того, щоб не ігнорувати цю частину запису, ту саму процедуру повторюють, починаючи з іншого кінця ряду. Таким чином будуть отримані $2N_s$ сегментів.

3. Обчислення локального тренду з використанням методу найменших квадратів. Потім визначається ряд

$$Y_s(i) = Y(i) - p_v(i), \quad (2.12)$$

де $p_v(i)$ є поліноміальною апроксимацією v -го сегменту.

4. Обчислення дисперсії для кожного з $2N_s$ сегментів

$$F_s^2(v) = \langle Y_s^2(i) \rangle = \frac{1}{s} \sum_{i=1}^s Y_s^2[(v-1)s + i]. \quad (2.13)$$

Останнім етапом є отримання флуктуаційної функції DFA шляхом усереднення по всіх сегментах і пошук квадратного кореня:

$$F(s) = \sqrt{\frac{1}{2N_s} \sum_{v=1}^{2N_s} F_S^2(v)}. \quad (2.14)$$

Відомо, що для самоподібних рядів даних $F(s)$ зростає зі збільшенням тривалості сегмента s за степеневим законом

$$F(s) \sim S^H. \quad (2.15)$$

Таким чином, флуктуаційні функції $F(s)$ зручно графічно представляти у подвійному логарифмічному масштабі як функції s , тоді для визначення H можна скористатися методом лінійної регресії.

Значення $H=0.5$ відповідає некорельованому ряду (або відсутності самоподібності в досліджуваному процесі), а $H>0.5$ вказує на тривало-залежний процес. Значення H , які є близькими до 1, показують високий ступінь самоподібності процесу, при цьому $H=1$ є межею стаціонарності.

2.3 Фрактальні параметри

Крім показника Херста існують інші фрактальні параметри, які дозволяють оцінювати степінь самоподібності часового ряду, а також пов'язані з ним властивості.

Фрактальна розмірність. Фрактальний об'єкт, з математичної точки зору, має дробову розмірність, яку можна обчислити за формулою:

$$D = \frac{\log N}{\log \frac{1}{r}}, \quad (2.16)$$

де N – кількість рівних частин, на які ділять об'єкт, r – коефіцієнт масштабу.

Фрактальна розмірність може розглядатись як міра нерівності порізаної поверхні об'єкту $D \in [n, n + 1)$ у n -мірному просторі, при чому більш нерівні поверхні відповідають високим значенням D .

Фрактальна розмірність виводиться через коефіцієнт Херста таким чином:

$$D = 2 - H. \quad (2.17)$$

Фрактальна розмірність узагальнює поняття евклідової геометричної розмірності i , на відміну від останньої, може приймати нецілочисельні значення. За допомогою даної розмірності можна класифікувати процеси за властивостями хаотичності або складності.

Кореляційний параметр. Розглянемо автокореляційну функцію. Нехай X_t значення випадкового процесу у момент часу t (t може бути дійсним числом, якщо процес безперервний, і цілим, якщо процес дискретний). Якщо X_t має середнє значення μ і дисперсію σ^2 , то автокореляція X_t визначається так

$$r(t - s) = r(t, s) = \frac{E[(X_t - \mu)(X_s - \mu)]}{\sigma^2}. \quad (2.18)$$

Випадковий процес X_t має повільно спадаючу залежність, якщо для його автокореляційної функції виконується умова

$$r(k) \sim k^{-\beta} L_1(k) + C_1, k \rightarrow \infty, \quad (2.19)$$

де $C_1 = const, 0 < \beta < 1$ – кореляційний параметр, L_1 – функція, яка повільно змінюється на нескінченності, $\lim_{t \rightarrow \infty} \frac{L_1(tx)}{L_1(t)} = 1, \forall x$.

Процеси з повільно спадаючою залежністю характеризуються автокореляційною функцією, яка спадає за степеневим законом при збільшенні часу затримки.

Зв'язок кореляційного параметру з показником Херста визначає формула:

$$\beta = 2(1 - H). \quad (2.20)$$

Спектральний показник. У частотній області повільно спадаюча залежність приводить до степеневому закону поведінки спектральної щільності процесу.

Нехай $r(t)$ – автокореляційна функція процесу X_t , тоді функція $S(f) = \int_{-\infty}^{\infty} r(\tau)e^{-i2\pi f\tau} d\tau$ є спектральною щільністю процесу X_t .

Випадковий процес X_t має повільно спадаючу залежність, якщо для спектральної щільності $S(f)$ виконується умова

$$S(f) \sim f^{-b} L_2(f) + C_2, \quad (2.21)$$

де $f \rightarrow 0$, $0 < b < 4$, L_2 – функція, яка повільно змінюється в нулі, $C_2 = const$.

Залежність спектрального показника від показника Херста:

$$b = 2H + 1. \quad (2.22)$$

Фрактальний показник. Фрактальний показник α – це міра того, наскільки гостроверхою є функція щільності ймовірності. У нормальному розподіленні (рис. 2.2а) $\alpha=2$. У випадку розподілень Парето (рис. 2.2б) $0 < \alpha < 2$. Даний показник характеризує степінь важкості «хвоста» розподілення (індекс «хвоста»).

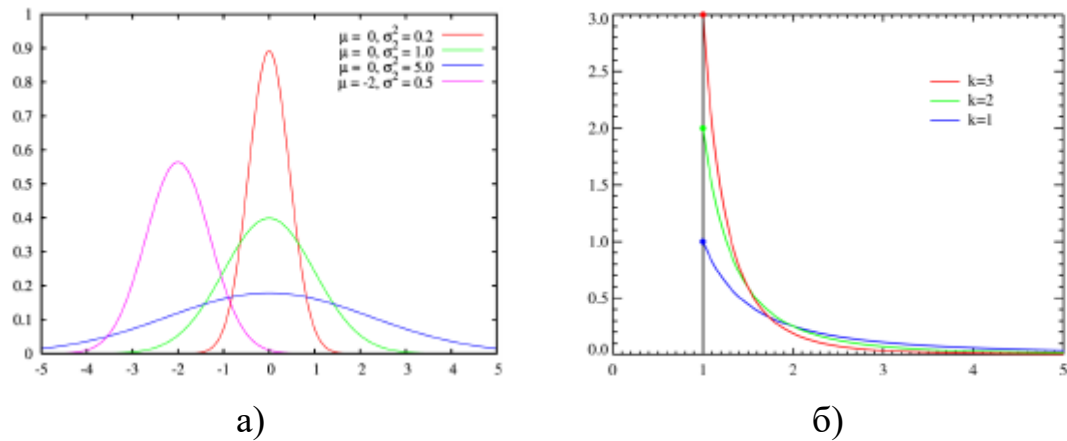


Рисунок 2.2 – Щільність розподілень: а – нормальне розподілення; б – розподілення Парето.

Випадкова величина ξ має розподілення з «важким хвостом», якщо:

$$P(\xi \geq x) = 1 - F(x) = x^{-\alpha} L_1(x), \quad (2.24)$$

де L_1 – функція, яка повільно змінюється на нескінченності, параметр $0 < \alpha < 2$ – індекс «хвоста».

Параметр α пов'язаний з параметром Херста таким співвідношенням:

$$\alpha = 3 - 2H. \quad (2.25)$$

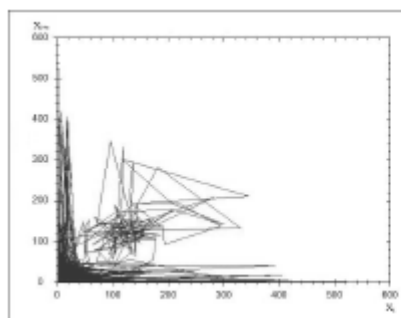
Оскільки всі описані фрактальні параметри можуть бути представлені через параметр Херста (рис. 2.3), то найчастіше саме його використовують для аналізу і дослідження мережевого трафіку.



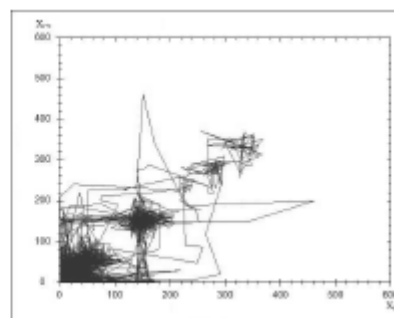
Рисунок 2.3 – Взаємозв'язок фрактальних параметрів.

2.4 Мережевий трафік як детермінований хаос

Властивості трафіку мереж передачі даних можна розглядати з позиції теорії детермінованого хаосу [33]. Ця теорія описує поведінку систем, чутливих до початкових умов. Виділення аттракторів на графіку фазового простору говорить про фазові перехідні процеси, що також може дозволити зробити прогноз поведінки системи. На рис. 2.4 показані приклади фазових діаграм для вхідного (рис. 2.4а) і вихідного (рис. 2.4б) мережевого трафіку.



а



б

Рисунок 2.4 – Фазові діаграми мережевого трафіку.

У якості параметру, який характеризує хаотичну поведінку систем, використовують експоненту Ляпунова [38]:

$$\lambda_1(i) = \frac{1}{i\Delta t} \frac{1}{(M-i)} \sum_{j=1}^{M-i} \ln \frac{d_j(i)}{d_j(0)}, \quad (2.26)$$

де Δt – період вибірки, $d_j(i)$ – відстань між j -ю парою найближчих сусідів після i дискретних кроків, M – кількість відновлених точок.

Для усіх типів трафіку старший показник Ляпунова λ_1 приймає значення від 1.2 до 2.5, що дозволяє працювати з трафіком з використанням методів нелінійної динаміки.

Показник Ляпунова для процесів розподіленої пам'яті серверу приймає значення нижче нуля, що характеризує високу степінь періодичності процесів, яка приймається як шум.

Мірою хаотичності процесу можна обрати величину кореляційної ентропії сигналу [34]. Для цього спочатку потрібно отримати кореляційний інтеграл (кількість пар точок не більша за r):

$$C(r) = \frac{2}{m(m-1)} \sum_{i=0}^{m-2} \sum_{j=i+1}^{m-1} \theta \left(r - p(x_i, x_j) \right), \quad (2.27)$$

де θ – функція Хевісайда ($\theta(\alpha) = 1$ при $\alpha \geq 0$, $\theta(\alpha) = 0$ при $\alpha < 0$), p – відстань у n -вимірному фазовому просторі, m – кількість точок x_i на аттракторі. Після чого розглядається залежність кореляційного інтегралу від r і розмірності фазового простору n :

$$C(r, n) \sim r^{D_2} \exp(-nK), \quad (2.28)$$

звідки отримується значення ентропії K :

$$K(r, n) = \ln \frac{c(r, n)}{c(r, n+1)}. \quad (2.29)$$

Чисельне значення кореляційної ентропії не тільки служить показником хаотичності процесу, але й показує оптимальний час, на який можна виконати прогнозування ряду.

Висновки за розділом 2

Задача моделювання і аналізу самоподібного трафіку є необхідною для ефективного управління мережами передачі даних та забезпечення якості їх обслуговування. Тому важливо обрати найбільш задовільну модель, яка б відтворювала поведінку характеристик комп'ютерних мереж.

У роботі пропонується використати потік FSNDP як модель трафіку комп'ютерної мережі. Проводити оцінку самоподібності моделі пропонується з використанням показнику Херста, оскільки всі інші показники можна звести до нього.

РОЗДІЛ 3. МОДЕЛЮВАННЯ І АНАЛІЗ ТРАФІКУ

3.1 Програмна реалізація моделі самоподібного трафіку

Для створення моделі самоподібного трафіку був обраний алгоритм процесу FSNDP. Блок-схема алгоритму моделювання показана на рис. 3.1.

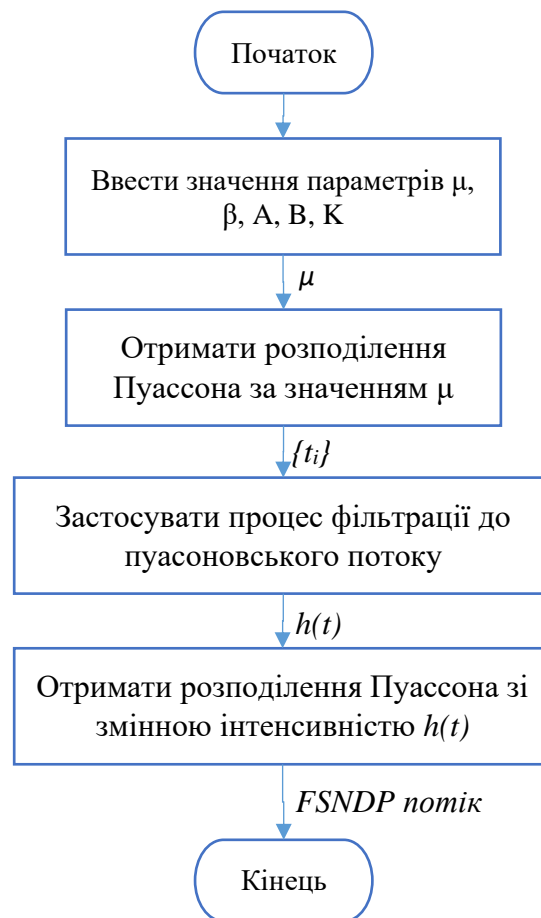


Рисунок 3.1 – Блок схема алгоритму FSNDP процесу.

Параметри μ , β , A , B , K однозначно визначають пуасоновський процес.

Програмна реалізація моделі самоподібного трафіку була розроблена за допомогою мови програмування Python. Були використані такі бібліотеки:

- numpy – бібліотека математичних функцій та роботи з масивами даних;
- random – бібліотека генерації випадкових чисел;
- matplotlib – бібліотека для побудови графіків різного типу (графіки, гістограми, діаграми на площині тощо);
- seaborn – бібліотека для візуалізації даних і виділення статистичних особливостей.

Спочатку генеруємо набір значень за розподіленням Пуассона:

```
t = np.random.poisson(mu, num) .
```

Далі передаємо отриману множину значень до функції фільтрації.

Отримуємо вихідні значення процесу FSNDP, подавши на вхід до другого пуассоновського потоку змінні інтенсивності.

На графіку (рис. 3.2) показаний приклад отриманого самоподібного трафіку з параметром Херста 0.7 ($\beta=0.8$).

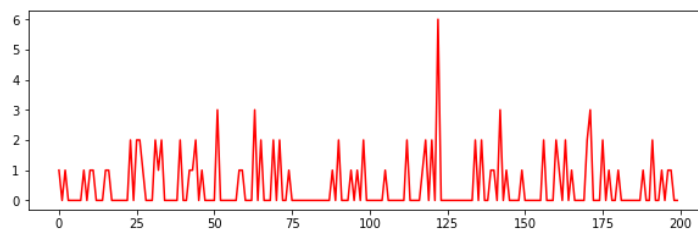


Рисунок 3.2 – Модель процесу FSNDP при $\beta=0.8$.

3.2 Аналіз реального трафіку мережі передачі даних

3.2.1 Опис набору даних

Для аналізу був обраний набір даних за результатом моніторингу роботи корпоративної мережі технічного університету. Були зібрані дані за такими характеристиками мережі:

- об'єм кешованої пам'яті;
- об'єм буферизованої пам'яті;
- час простою процесору під час операцій вводу/виводу;
- час процесора у режимі очікування;
- час обробки процесором користувацьких процесів;
- час обробки процесором системних процесів;
- об'єм вільної пам'яті;
- вхідний/вихідний трафік (біт/с);
- кількість процесорів ОС;
- кількість процесорів web-серверу;
- сумарне навантаження процесору.

Дані були отримані за різні часові проміжки для різних параметрів, від 1 до 60 секунд протягом доби. Часова залежність об'єму трафіку в одиницю часу показана на рис. 3.3 (а, б) для вхідного і вихідного трафіків. На рис. 3.4 (а,б) показана залежність для завантаження процесору і об'єму вільної пам'яті відповідно, а на рис. 3.5 (а, б) – залежність для кількості процесів і об'єму кешованої пам'яті.

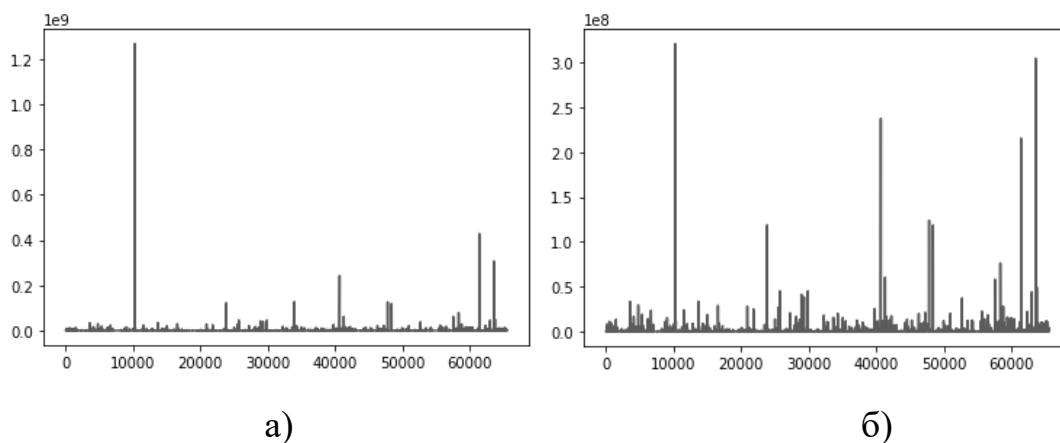


Рисунок 3.3 – Часова залежність об'єму трафіку: а – вхідний трафік; б – вихідний трафік.

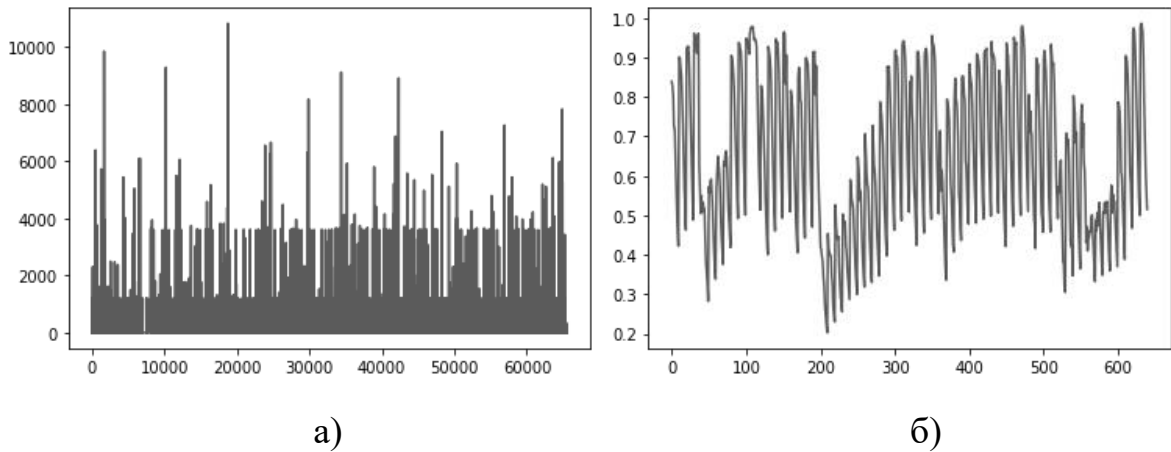


Рисунок 3.5 – Часова залежність завантаження ЦПУ (а) та вільної пам'яті (б).

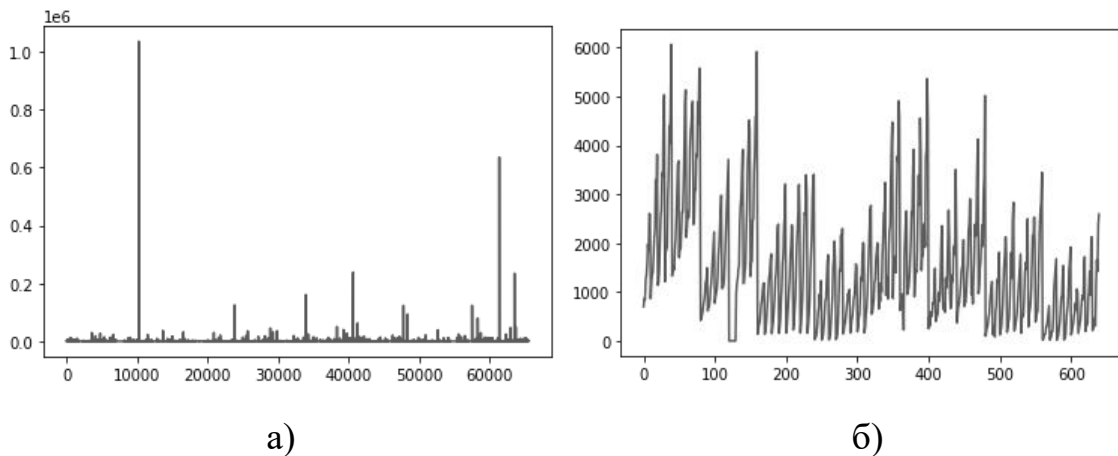


Рисунок 3.6 – Часова залежність для кількості процесів (а) і об'єму кешованої пам'яті (б).

3.2.2 Статистичний аналіз параметрів трафіку

Для визначення особливостей досліджуваних процесів варто провести аналіз даних моніторингу мережі засобами статистичного аналізу.

Аналіз функції розподілення. Щільність розподілення ймовірностей випадкової величини X – це перша похідна від інтегральної функції розподілення ймовірностей $F(x)$:

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = F'(x) = \frac{dF(x)}{dx}, \quad (3.1)$$

оскільки ймовірність того, що випадкова величина X міститься на проміжку $[x, x + \Delta x]$:

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x) \approx dF(x) = f(x)dx. \quad (3.2)$$

Графічно щільність розподілення ряду можна оцінити за гистограмою частот виникнення проміжків даних.

На рис. 3.7 (а, б) показані гистограми за вихідним трафіком (а) та вихідним трафіком (б), на рис. 3.8 – гистограми буферизованої пам'яті (а) і кешованої пам'яті (б), на рис. 3.9 – гистограми часу очікування процесору (а) і часу очікування у режимі вводу/виводу (б), на рис. 3.10 – гистограми вільної пам'яті (а) і сумарного навантаження процесору (б).

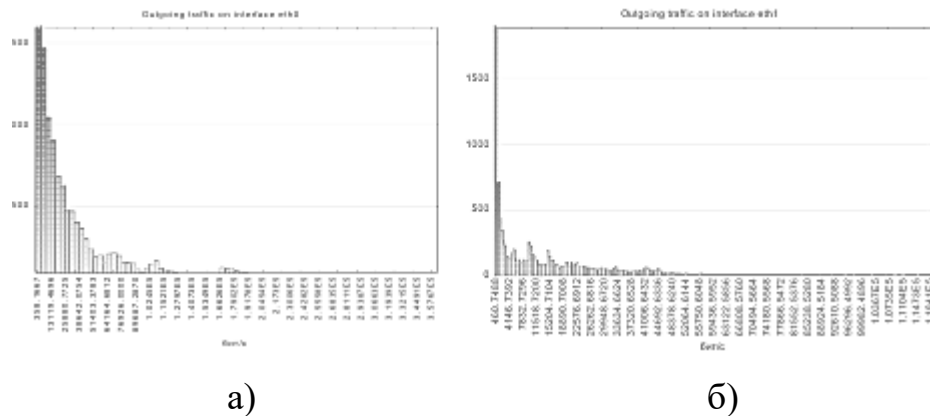


Рисунок 3.7 – Гистограми частот за вихідним трафіком (а) та вихідним трафіком (б).

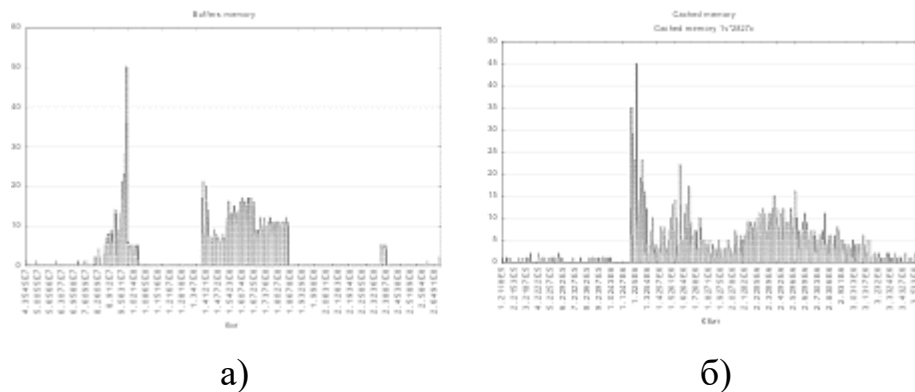


Рисунок 3.8 – Гистограми буферизованої пам'яті (а) і кешованої пам'яті (б).

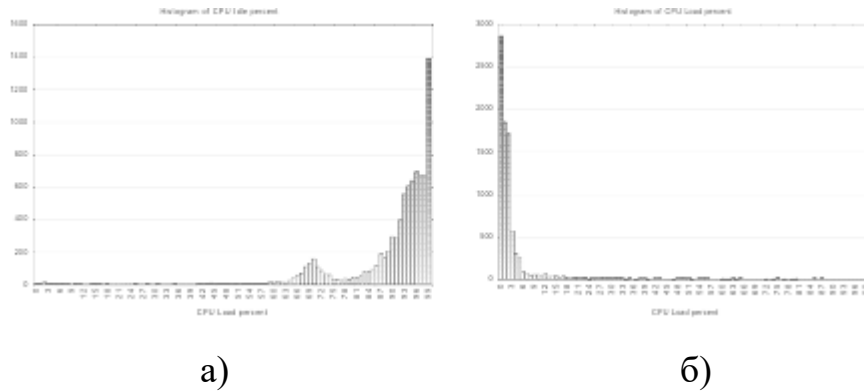


Рисунок 3.9 – Гістограми часу очікування процесору (а) і часу очікування у режимі вводу/виводу (б).

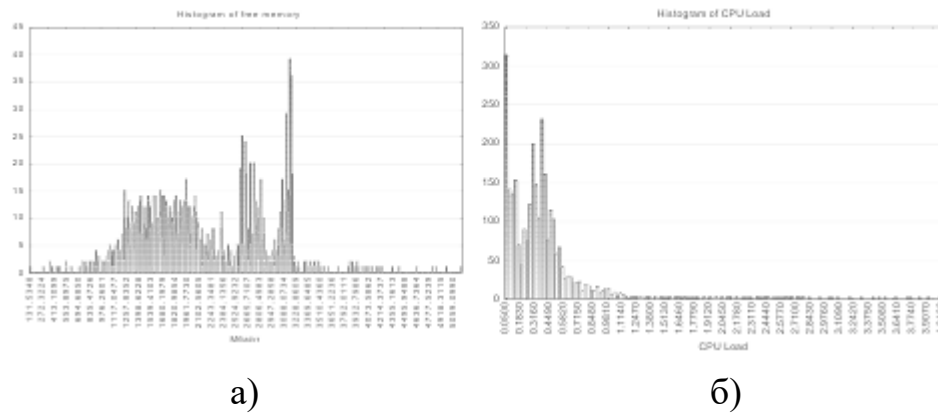


Рисунок 3.10 – Гістограми вільної пам'яті (а) і сумарного навантаження процесору (б).

З гістограм видно, що щільність розподілення ймовірності процесів передачі вхідного і вихідного трафіків має степеневий характер і, відповідно підкоряються деякому класу розподілень з важким хвостом (Стюдента або Парето). Параметри апаратних ресурсів серверу можна описати або степеневою функцією щільності розподілення, або відповідають нормальному розподіленню.

Аналіз функції автокореляції. Нехай заданий часовий ряд, складений із рівновіддалених за часом значень $(x_i), i = 1, \dots, N$. Необхідно виділити кореляції між значеннями x_i і x_{i+s} для різних значень s . Спершу обчислюється середнє значення $\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i$. Чисельно величина кореляції між значеннями \tilde{x} ,

віддаленими на s кроків, визначається за допомогою функції автоковаріації $C(s) = \langle \tilde{x}_i, \tilde{x}_{i+s} \rangle$ або автокореляції $C(s) / \langle \tilde{x}_i^2 \rangle$.

Якщо величина $C(s)$ убуває експоненційно, то (x_i) має короткострокову залежність і $C(s) \sim \exp\left(\frac{-s}{t_x}\right)$, а якщо $C(s)$ убуває за степеневим законом $C(s) \propto s^{-\gamma}$ з показником кореляції $0 < \gamma < 1$, то процес має довгострокову кореляцію. Нестационарність часового ряду у такому випадку робить важким пошук середнього значення.

На рис. 3.11 показані графіки кореляційних функцій вхідного трафіку (а) і вихідного трафіку (б), на рис. 3.12 – графіки кореляційних залежностей сумарного навантаження процесору (а) і вільної оперативної пам'яті (б).

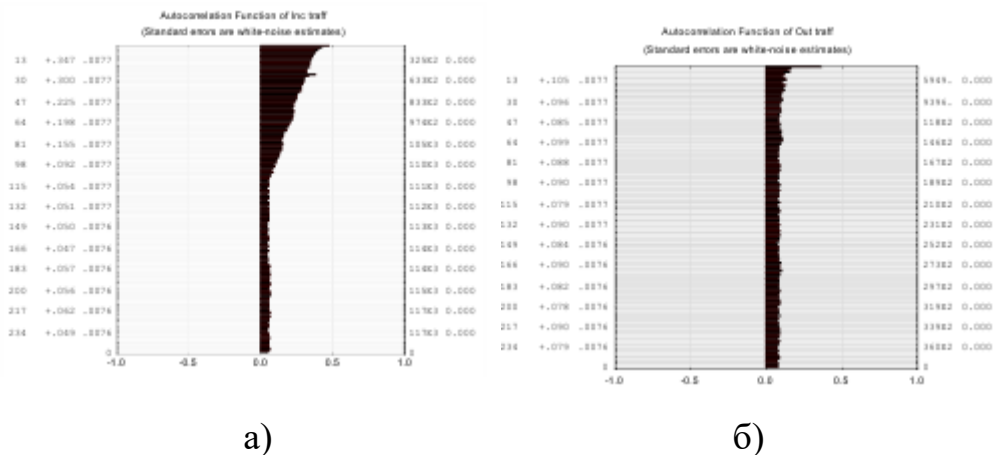


Рисунок 3.11 – Графіки АФК для вхідного трафіку (а) і вихідного трафіку (б).

З графіків (рис. 3.11 і 3.12) можна зробити висновок, що процес передачі трафіку має повільно згасаючу залежність. Також варто відмітити, що АКФ для процесу виділення пам'яті має чітку періодичність. Отже, у даному випадку процес характеризується логікою роботи операційної системи, що привносить сильну періодичну складову.

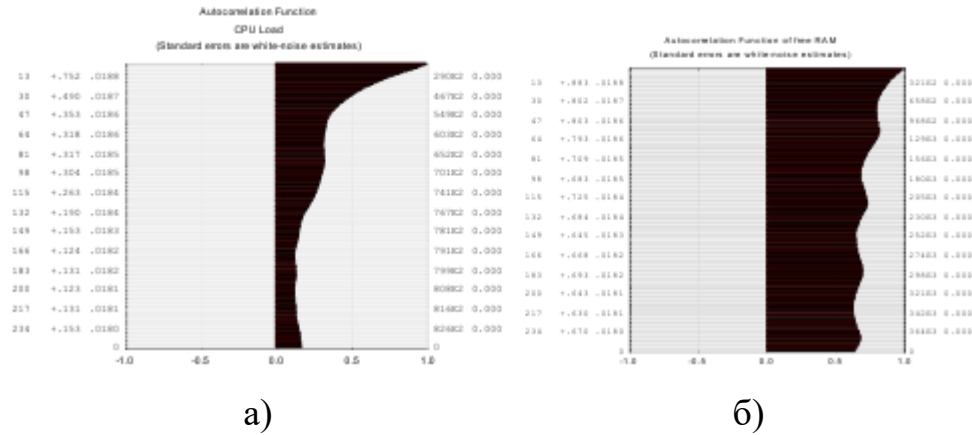


Рисунок 3.12 – Графіки АФК для сумарного навантаження процесору (а) і вільної оперативної пам'яті (б).

Спектральний аналіз – це інструмент обробки даних, який має в основі різні інтегральні перетворення. Спектром $X(i)$ називають деяку функцію $F(w)$, отриману у відповідності до визначеного алгоритму. Прикладами спектрів є Фур'є перетворення і вейвлет-перетворення.

Перетворення Фур'є перетворює функцію у сукупність її частотних складників, тобто це інтегральне перетворення, яке розкладає вхідну функцію за базисними синусоїдальними функціями:

$$F(w) = \int_{-\infty}^{\infty} y(x) \exp(-iwx) dx. \quad (3.3)$$

Дискретне перетворення Фур'є має вигляд:

$$f(t) = \sum_{n=-\infty}^{\infty} C_n \exp(i\beta w_n t), \quad (3.4)$$

де $w_n = nw_0 = n(2\pi/T)$ – кругова частота n -ї гармонічної складової, C_n – комплексна амплітуда n -ї гармоніки:

$$C_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) \exp(iw_n t) dt. \quad (3.5)$$

Сукупність C_n є спектр функції $f(t)$. На рис. 3.13 – 3.16 показані графіки спектральної щільності від частоти для основних апаратних характеристик серверу, а також мережевого навантаження.

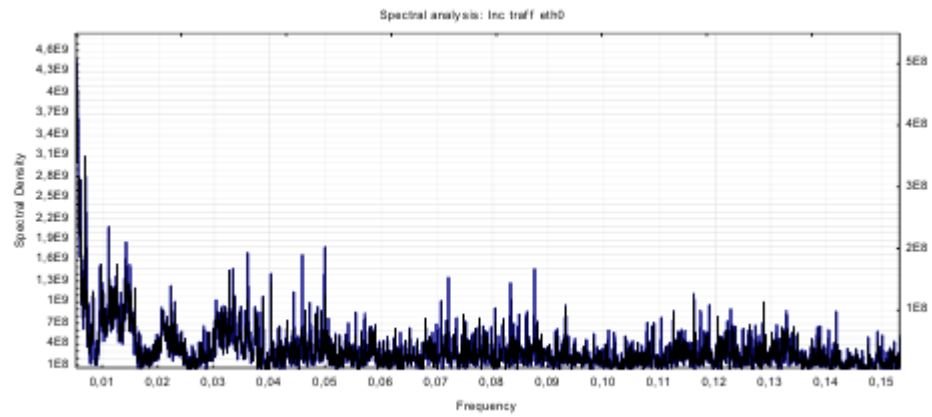


Рисунок 3.13 – Спектр вхідного трафіку.

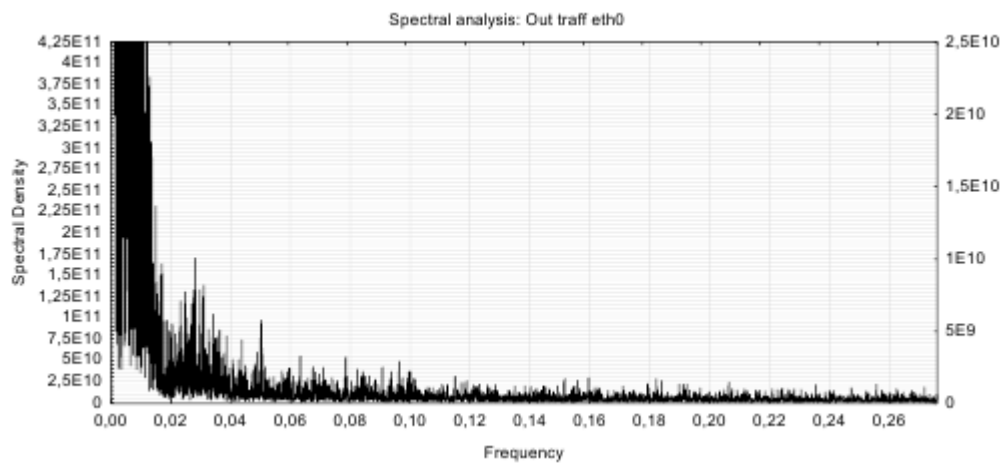


Рисунок 3.14 – Спектр вихідного трафіку.

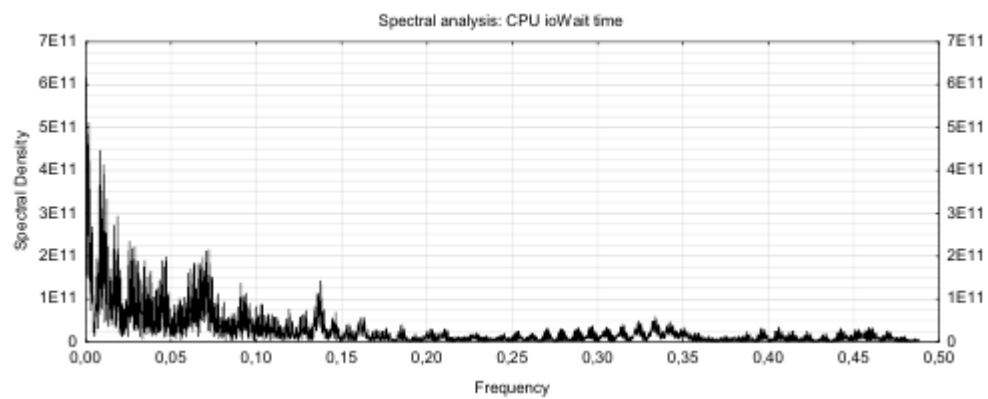


Рисунок 3.15 – Спектр часу перебування процесору в режимі очікування.

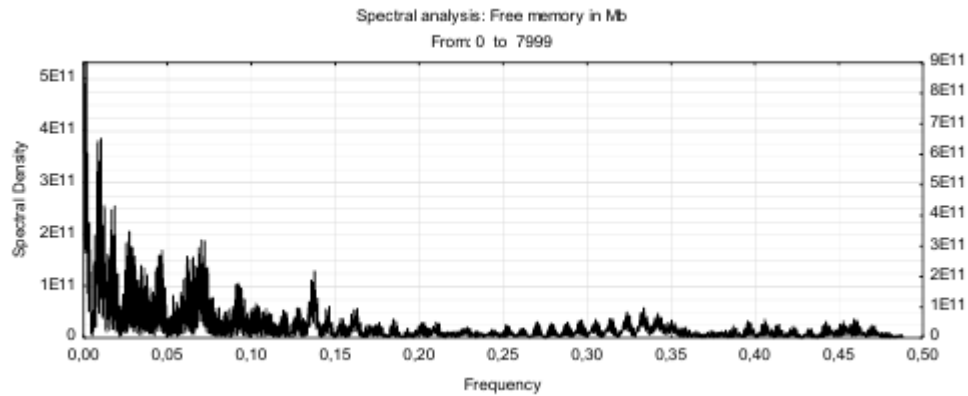


Рисунок 3.16 – Спектр вільної оперативної пам'яті.

Кореляційний і регресійний аналіз. У табл. 3.1 показані значення парної кореляції між трафіком і апаратним навантаженням. Виділені напівжирним найбільш значущі величини.

Таблиця 3.1

Коефіцієнти кореляції параметрів

	Вхідний трафік	Вихідний трафік	Сумарне навантаж.	Об'єм вільної пам'яті	Кількість користув. процесів	Час очікування ЦПУ
Вхідний трафік	1	0.23	0.17	-0.14	0.2	-0.12
Вихідний трафік	0.23	1	0.12	-0.39	0.27	-0.1
Сумарне навантаж.	0.17	0.12	1	-0.24	0.39	-0.7
Об'єм вільної пам'яті	-0.14	-0.39	-0.24	1	-0.33	0.35
Кількість користув. процесів	0.2	0.27	0.39	-0.33	1	-0.04
Час очікування ЦПУ	-0.12	-0.1	-0.7	0.35	-0.04	1

Усі значення парної кореляції легко інтерпретуються. Зростанню мережевого навантаження відповідає зростання навантаження на апаратні

ресурси серверу. У першу чергу це стосується сумарного навантаження і виділення оперативної пам'яті. Вихідний трафік сильно корелює з кількістю користувацьких процесів, що можна пояснити тим, що джерелом вихідного трафіку є локальні процеси серверу.

3.2.3 Аналіз нелінійних властивостей

Під аналізом нелінійних властивостей трафіку мається визначення степені самоподібності за допомогою показнику Херста.

Параметр Херста може бути оцінкою довгострокової залежності часового ряду. Оцінка параметру допоможе зробити висновок про самоподібність процесу і дозволить в подальшому застосувати до нього ряд математичних методів з прогнозування фрактальних процесів. Виконання короткострокових прогнозів навантаження серверу може допомогти в розробці методів апаратної і програмної оптимізації мережі.

Параметр Херста H для розглядуваного набору даних був оцінений за допомогою R/S аналізу виборки. У таблиці 3.2 наведені значення показнику Херста для мережевого трафіку і основних апаратних характеристик. Як і очікувалось, значення параметру показало наявність довгострокової залежності і самоподібності.

Таблиця 3.2

Показники Херста для параметрів мережі

Характеристика	Параметр Херста
Об'єм буферизованої пам'яті	0.9656
Об'єм кешованої пам'яті	0.9868
Час процесору в режимі очікування	0.9575
Час обробки системних задач	0.9903
Об'єм вільної пам'яті	0.9336
Вхідний трафік	0.9775
Вихідний трафік	0.9712
Кількість процесів ОС	0.8835

Також підтверджена гіпотеза, яка вказує зв'язок між параметром Херста і інтенсивністю трафіка (табл. 3.3).

Таблиця 3.3

Залежність показнику Херста від інтенсивності трафіку

	Значення параметру Херста			
	00:00 – 6:00	6:00 – 12:00	12:00 – 18:00	18:00 – 24:00
Вхідний трафік	0.714	0.6154	0.9836	0.9516
Вихідний трафік	0.8180	0.7385	0.9766	0.9112

Висновки за розділом 3

У розділі наведена програмна реалізація процесу моделювання самоподібного трафіку для різних значень коефіцієнту самоподібності.

Також проведений аналіз статистичних характеристик процесів, які відбуваються у реальній роботі корпоративної мережі. Аналіз щільностей розподілення ілюструє наявність довготривалої залежності, а також про наявність розподілення з довгим хвостом.

Методами кореляційного аналізу були визначені основні апаратні параметри, які чуттєві до підвищення інтенсивності трафіку.

До розглядуваного набору даних був обчислений параметр Херста, значення якого підтвердило самоподібність трафіку мережі та її процесів.

ВИСНОВКИ

Розвиток сучасних локальних і глобальних мереж є інтенсивним, розширюються апаратні можливості мережевого обладнання, зростає кількість користувачів й загальний об'єм трафіку. Тому постають проблеми підвищення вимог до мережевого і серверного обладнання, підтримки необхідного рівню якості обслуговування. Розвиток мережевого обладнання спирається на адекватні математичні моделі параметрів трафіку й інструменти моделювання мережевих процесів. Характер трафіку мереж визначається множиною факторів – від поведінки користувачів або програмного забезпечення, до протоколів передачі й обладнання, яке використовується. Тому, дослідження і моделювання основних характеристик трафіку та його поведінки є важливою задачею.

У роботі проведений аналіз існуючих моделей трафіку комп'ютерних мереж. У першому розділі були розглянуті такі моделі трафіку як: моделі, засновані на процесах відновлення, моделі, засновані на марківських процесах, авторегресійні моделі, самоподібні моделі та моделі трафіку додатків. Для подальшого моделювання були обрані самоподібні моделі.

У другому розділі був проаналізований процес моделювання самоподібного трафіку, розглянуті алгоритми моделювання, зокрема фрактальний дробовий процес, який дозволяє врахувати властивості самоподібності. Був розглянутий показник Херста як основний показник нелінійності та самоподібності трафіку і методи його оцінки. Також звернута увага на можливість розглядати трафік з точки зору нелінійної динаміки як детермінований хаос.

Третій розділ роботи присвячений розробці і опису програмної моделі самоподібного трафіку комп'ютерної мережі й її тестуванню. Крім того був розглянутий трафік реальної комп'ютерної мережі та проведений його статистичний аналіз і аналіз нелінійних властивостей характеристик.

Для розглядуваного прикладу трафіку реальної корпоративної мережі був визначений самоподібний характер процесів. Отже отримана програмна модель самоподібного трафіку може бути використана у практичній діяльності адміністраторів та проектувальників комп'ютерних мереж для визначення параметрів мережевого обладнання.

Таким чином усі задачі дослідження були виконані і мета дослідження досягнута.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Fishman, G.S. Principles of Discrete Event Simulation / G.S. Fishman. – 1978.
2. Frost, V. Traffic modeling for telecommunications networks / V. Frost, B. Melamed // IEEE Communications Magazine. – 1994.
3. Abdelnaser, A. Traffic Models in Broadband Telecommunication Networks / A. Abdelnaser // Department of Electrical Engineering. – 1996.
4. Chen, T.M. The Handbook of Computer Networks / T.M. Chen // Southern Methodist University. – 2007.
5. Brandauer, C. Comparison of Tail Drop and Active Queue Management Performance for built-data and Web-like Internet Traffic / C. Brandauer, C. Diot. – 2001.
6. Chen, T. Network Traffic Modelling / T. M. Chen // Wiley. – 2007.
7. Hefles, H. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance / H. Hefles, D. Lucantoni // IEEE Journal on Selected Areas in Communications. – 1986.
8. Abdelnaser, A. Traffic Models in Broadband Networks / A. Abdelnaser // IEEE Communications Magazine. – 1997.
9. Reichl, P. A Generalized TES Model for Periodical Traffic / P. Reichl // IEEE International Conference. – 1998.
10. Land, W. On the self-similar nature of Ethernet traffic / W.Land, M.Taqqu, W.Willinger // IEEE/ACM Transactions on Networking. – 1994.
11. Crovella, M. Self-similarity in World Wide Web traffic: evidence and possible causes / M. Crovella, A. Bestavros // IEEE/ACM Transactions on Networking. - 1997.
12. Роздымаха Е. А. Прогнозирование трафика сети мобильной связи третьего поколения / Е. А. Роздымаха, А. В. Омельченко, А. В. Федоров // Problems of Infocommunications Science and Technology: conference proceedings of

the 2015 Second International Scientific-Practical Conference Proceedings (PICS&T 2015). – Харків, 2015. – С. 54–57.

13. Игнатенко Е. Г. Методика краткосрочного прогнозирования трафика телекоммуникационных сетей / Е. Г. Игнатенко, И. В. Дегтяренко, Н. В. Червинская, И. Н. Яремко // Збірник наукових праць ДонІЗТ: Зб. наук. пр. – Донецьк: ДонІЗТ, 2011. – № 28. – С. 102–107.

14. Крюков Ю. А. ARIMA – модель прогнозирования значений трафика / Ю. А. Крюков, Д. В. Чернягин // Т-Comm. – 2013. – № 7. – С. 97–98

15. Покровская М. А. Сравнительный анализ методов прогнозирования трафика мультимедийных приложений / М. А. Покровская // Информационные технологии и вычислительные системы. – 2011. – № 2. – С. 41–49.

16. Fraleigh, C. Packet-level traffic measurements from the Sprint IP backbone / C. Fraleigh, S. Moon // IEEE Network. – 2003.

17. Meiss, M. On the lack of typical behavior in the global Web traffic network / M. Meiss, F. Menczer // 14th International World Wide Web Conf. – 2005.

18. Li, M. Characteristics of streaming media stored on the Web / M. Li, M. Claypool // ACM Transactions on Internet Technology. – 2005 .

19. Choi, H-K. A behavioral model of Web traffic / H-K. Choi, J. Limb // In proc. of 7th International Conf. on Network Protocols (ICNP'99). – 1999.

20. Meiss, M. On the lack of typical behavior in the global Web traffic network / M. Meiss, F. Menczer, A. Vespignani // In proc. of 14th International World Wide Web Conf. – 2005.

21. Saroiu, S. An analysis of Internet content delivery systems / S. Saroiu, K. Gummadi, R. Dunn, S. Gribble, H. Levy // ACM SIGOPS Operating Systems Review. – 2002.

22. Sen, S. Analyzing peer-to-peer traffic across large networks / S. Sen, J. Wang // IEEE/ACM Transactions on Networking. – 2004.

23. Strelkovskaya I. Spline approximationbased restoration for self-similar traffic / I. Strelkovskaya, I. Solovskaya, N. Severin, S. Paskalenko // Eastern-European Journal of Enterprise Technologies. – 2017. – № 3/4 (87). – P. 45–50.

24. Комплексна оцінка характеристик та ідентифікація трафіку в інформаційних телекомунікаційних мережах / Л. В. Будкова, В. І. Корнієнко // Системи обробки інформації. – 2013. – № 2 (109). – С. 207–211.

25. Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes / M. E. Crovella, A. Bestavros // IEEE Transactions on Networking. – 1997. – Vol. 5, № 6. – P. 835–846.

26. Моделювання самоподібного трафіка в інформаційних телекомунікаційних мережах / Л. В. Будкова, В. І. Корнієнко // Вісник КрНУ імені Михайла Остроградського. – 2013. – Вип. 4. – С. 46-52.

27. Водопьянов С. В. Применение моделей трафика данных для мониторинга компьютерных сетей системы управления воздушным движением / С. В. Водопьянов, В. И. Дрововозов // Проблеми інформатизації та управління. – 2012. – № 1 (37). – С. 30-35.

28. Карпухин А. В. Применение методов нелинейной динамики и фрактального анализа для оценивания работы инфокоммуникационных систем с протоколом ТРС / А. В. Карпухин, Л. О. Кириченко, Д. И. Гривцев, А. А. Ткаченко // Cloud of Science. – 2014. – № 2. – С. 258-271.

29. Стрелковська І. В. Прогнозування характеристик самоподібного трафіку за допомогою сплайн-екстраполяції / І. В. Стрелковська, І. М. Соловська, А. О. Макоганюк, М. В. Северин // Вісник Університету «Україна». – 2019. – №1 (22). – С. 188-195.

30. Cortez P. Multiscale Internet traffic forecasting using neural networks and time series methods / P. Cortez, M. Rio, M. Rocha, P. Sousa // Expert Systems. – 2012. – Issue 29 (2). – P. 143-155.

31. Макаришкін Д. А. Аналіз особливостей статистичних властивостей трафіку сучасних інформаційних мереж та підходів до моделювання інформаційних потоків / Д. А. Макаришкін, Л. О. Ковтун, В. М. Ткачук, М. О. Кінаш // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2019. – Вип. 1. – С. 62-67.

32. Paxson V. Wide-area traffic: the failure of Poisson modeling / V. Paxson, S. Floyd // ACM SIGCOMM Computer Communication Review / ACM. – 1994. – Vol. 24. – P. 257-268.

33. Park K. Self-similar network traffic: An overview / K. Park, W. Willinger // Self-similar network traffic and performance evaluation. – 2000. – P. 1–38.

34. Колесніков А. В. Нелінійно-динамічні моделі мережевого трафіку / А. В. Колесніков, І. П. Іванов, М. А. Басараб // Нелінійний світ. – 2014. – Т. 12, № 4. – С. 44–56.

35. Матичин І. І. Моделювання та аналіз трафіку в телекомунікаційних системах та мережах / І. І. Матичин, В. В. Онищенко // Вісник ДУІКТ. – 2013. – № 4. – С. 20-27.

36. Leland W. E. On the Self-Similar Nature of Ethernet Traffic (Extended Version) / W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson // IEEE/ACM Trans, on Networking. – 1994. – Vol.2, Issue 1. – P. 1-15.

37. Никульчев Е. В. Вычисление характеристик динамического хаоса по трафику компьютерных сетей / Е. В. Никульчев, С. В. Паяин, Д. А. Питиков, Е. В. Плужник // Фундаментальные исследования. – 2014. – № 8. – С. 812-816.

38. Xiong W. Anomaly secure detection methods by analyzing dynamic characteristics of the network traffic in cloud communications / W. Xiong, H. Hu, N. Xiong, L. T. Yang, W. C. Peng, X. Wang, Y. Qu // Information Sciences. – 2014. – Vol. 258. – P. 403–415.

39. Карпухин А. В. Моделирование хаотических явлений в инфокоммуникационных сетях: синергетический подход / А. В. Карпухин, А. А. Ткаченко // Проблемы телекоммуникаций. – 2013. – № 3 (12). – С. 36–52.

40. Ryu Bo. Modeling, Analysis and Simulation of Self-Similar Traffic using the Fractal-Shot-Noise-Driven Poisson Process / Bo Ryu, S. B. Lowen // Internet Traffic Modelling. – 1995.

41. Veres B. M. The chaotic nature of TCP congestion control / B. M. Veres // IEEE INFOCOM'2000. – 2000.

42. Нейронне вейвлет прогнозування та ідентифікація складних сигналів та об'єктів керування / В. І. Корнієнко, Г. В. Кузнецов, І. В. Гарнак // Інформаційні технології та комп'ютерна інженерія. – 2009. – № 2 (15). – С. 38–44.

ДОДАТКИ

Додаток А

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

Факультет комп'ютерних наук
Кафедра теоретичної та прикладної системотехніки
Освітньо-кваліфікаційний рівень Магістр
Галузь знань 12 – Інформаційні технології
Спеціальність 123 – Комп'ютерна інженерія

ЗАТВЕРДЖУЮ

Завідувач кафедри теоретичної та
прикладної системотехніки



д.т.н., проф. Шматков С. І.
«10» грудня 2022 року

З А В Д А Н Н Я **НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТА**

Заца Олександра Дмитровича

(прізвище, ім'я, по батькові)

1. Тема роботи: Моделі трафіку комп'ютерних мереж

керівник роботи: Шматков Сергій Ігорович, д.т.н., професор

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від “ ___ ” _____ 2022 року

№ _____

2. Строк подання студентом роботи 30 листопада 2022 р.

3. Перелік питань, які потрібно розробити:

3.1. Аналіз методів і технологій моделювання трафіку комп'ютерних мереж.

3.2. Розробка математичних моделей трафіку комп'ютерних мереж.

3.3. Програмна реалізація моделей трафіку комп'ютерних мереж, їх тестування та дослідження.

4. План роботи

№ з/п	Назва етапів дипломної роботи	Строк виконання етапів роботи
4.1	Формулювання цілей роботи та постановка задачі.	11.12.2021 – 01.01.2022
4.2	Аналіз існуючих технологій та методів моделювання трафіку комп'ютерних мереж.	02.01.2021 – 05.02.2022
4.3	Збір вихідних даних для моделювання трафіку науково-освітніх комп'ютерних мереж.	06.02.2022 – 15.04.2022
4.4	Розробка моделі трафіку комп'ютерних мереж із урахуванням самоподібності.	16.04.2022 – 30.05.2022
4.5	Програмна реалізація моделі трафіку комп'ютерних мереж із урахуванням самоподібності.	01.06.2022 – 02.09.2022
4.6	Аналіз результатів моделювання трафіку за критеріями самоподібності	03.09.2022 – 15.10.2022
4.7	Тестування розроблених моделей та їх дослідження.	16.10.2022 – 31.10.2022
4.8	Оформлення пояснювальної записки.	01.11.2022 – 30.11.2022

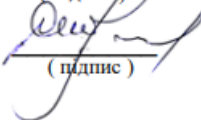
5. Дата видачі завдання 10.12.2021

Студент

Керівник роботи



(підпис)



(підпис)

Зац О.Д.

(прізвище та ініціали)

Шматков С.І.

(прізвище та ініціали)

Додаток Б

«Затверджую»

« ___ » _____ 2022 р.

ІНДИВІДУАЛЬНЕ ТЕХНІЧНЕ ЗАВДАННЯ


на розробку програмного виробу «Програмна модель трафіку мережі
передачі даних»

1.	Введення	Назва: Програмна модель трафіку мережі передачі даних. Область застосування: моделювання й аналіз параметрів комп'ютерних мереж.
2.	Підстава для розробки	1) Навчальний план ФКН за спеціальністю 123 – Комп'ютерна інженерія. 2) Завдання на дипломну роботу магістра № ___ від _____ 2021 р. (представити як Додаток А до пояснювальної записки до дипломної роботи);
3.	Призначення та цілі розробки	1) Модель призначена для використання в науково-дослідній роботі. 2) Цілі розробки: створити програмну модель трафіку комп'ютерної мережі.
4.	Початкові дані для розробки	В якості початкових даних для розробки використовується інформація про параметри комп'ютерної мережі.
5.	Вимоги до програмного продукту (моделі)	Програма повинна: 1) Представляти с себе програмну реалізацію моделі трафіку комп'ютерної мережі. 2) Забезпечувати можливість введення даних і отримання результатів роботи програми у вигляді рекомендацій, коефіцієнтів, таблиць тощо. 3) Забезпечувати автоматичний контроль введених даних і захист від некоректних дій користувача. 4) Для виконання програми повинен використовуватися ПК за будь-якою ОС сімейства Windows, Linux, Mac OS. 5) Вимоги до інформаційної та програмної сумісності: Windows версії 10 і вище, Linux/Unix (будь-який дистрибутив), Mac OS. 6) Вимоги до маркування та упаковки (не надаються). 7) Вимоги до транспортування і зберігання (не надаються). 8) Спеціальні вимоги (не надаються).

6.	Вимоги до програмної документації	Програмною документацією щодо розроблюваного програмного продукту вважати: <ol style="list-style-type: none"> 1) Справжнє технічне завдання на розробку програми (представити як Додаток Б до пояснювальної записки до дипломної роботи); 2) Опис програми (представити в Розділі 3 пояснювальної записки до дипломної роботи). 3) Частина тексту програми (представити як Додаток Г пояснювальної записки до дипломної роботи). 	
7.	Вимоги до техніко-економічних показників	<ol style="list-style-type: none"> 1) Оцінка економічної ефективності – непотрібна. 2) Визначення економічних переваг моделі в порівнянні з вітчизняними та зарубіжними аналогами – непотрібно. 	
8.	Стадії і етапи розробки	Дата	Назва етапу
		від 1 вересня 2022 до 14 вересня 2022	Аналіз технологій для розробки програмної моделі трафіку.
		від 15 вересня 2022 до 25 вересня 2022	Вибір параметрів комп'ютерної мережі для моделювання її трафіку. Пошук даних трафіку реальної мережі для аналізу.
		від 26 вересня 2022 до 10 жовтня 2022	Розробка програмної моделі трафіку комп'ютерної мережі.
		від 11 жовтня 2022 до 25 жовтня 2022	Тестування програмної моделі.
		від 26 жовтня 2022 до 20 листопада 2022	Аналіз реального трафіку мережі передачі даних
9.	Порядок контролю і приймання програмного продукту (моделі)	<ol style="list-style-type: none"> 1) Перевірку ходу розробки програми виконувати керівнику дипломної роботи раз в 3 тижні. 2) Випробування програмного продукту провести відповідно до програми та методики випробувань на базі комп'ютерного класу. 3) Захист розробленої моделі провести на засіданні Атестаційної комісії. 4) Пояснювальну записку подати на паперових носіях в 1 примірнику і в електронному вигляді в 1 примірнику на CD-R компакт-диску. 	

Виконавець

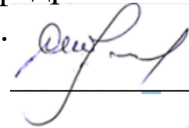
студент групи КІ-61

Зац О. Д. 

Замовник

доктор технічних наук,
професор,

завідувач кафедри ТПС

Шматков С.І. 

**Програма і методика випробувань
програмного виробу
«Програмна модель трафіку мереж передачі даних»**

1. Об'єкт випробувань

Об'єктом випробування є процес моделювання трафіку мереж передачі даних (розділ 1 ТЗ).

2. Мета випробувань

Перевірка відповідності функціональності розробленого методу заявленим функціональним можливостям в технічному завданні (Додаток Б).

3. Вимоги до програми або програмного виробу

Програмний виріб повинен надавати такі можливості:

- 1) можливість задавати параметри трафіку;
- 2) можливість проаналізувати показники самоподібності трафіку;
- 3) провести статистичний аналіз характеристик трафіку.

4. Вимоги до програмної документації

Склад програмної документації, що подається на випробування, включає:

- 1) Технічне завдання на розробку програмного виробу (представлено в Додатку Б до пояснювальної записки до дипломної роботи).
- 2) Ця програма і методика випробувань розробленого програмного виробу (представлена в Додатку В до пояснювальної записки до дипломної роботи).
- 3) Опис програмного виробу (представлено в розділі 3 пояснювальної записки до дипломної роботи).

4) Фрагмент програми (представлений в Додатку Г до пояснювальної записки до дипломної роботи)

5. Засоби і порядок випробувань

5.1. Засоби випробувань

1) Програмний виріб вимагає наявності графічного пришивидшувача для успішного запуску.

2) Для виконання програми потрібно, щоб на ПК було встановлено останню версію Python та доступ до бібліотек numpy, matplotlib, seaborn.

5.2. Порядок проведення випробувань

5.2.1. Перевірка програмної документації

1) Перевірка складу програмної документації. Перевірку здійснювати за критерієм наявності, представленої в ТЗ документації.

2) Перевірка якості програмної документації. Перевірку здійснювати за критерієм відповідності вимогам ГОСТ 19.301-79 ЕСПД. «Програма і методика випробувань».

5.2.2. Перевірка працездатності програмної моделі

Тест 1

1. Перевірку здійснюємо шляхом запуску програми, розміщеній у «main.py». Вводимо значення параметрів процесу моделювання трафіку. Отримуємо графік трафіку мережі.

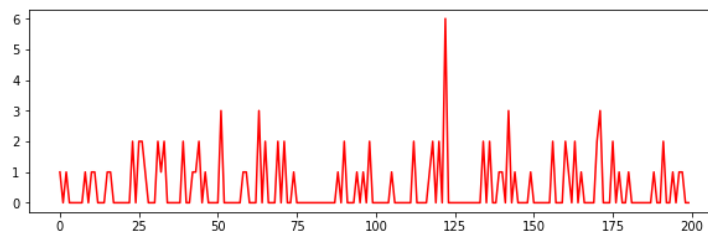


Рис. В.1 – Результат моделювання трафіку.

Тест 2

2. Перевірку здійснюємо шляхом запуску програми, розміщеній у «statistic.py», перевіряємо роботу програми для аналізу трафіку. Отримуємо гістограми розподілень характеристик мережі.

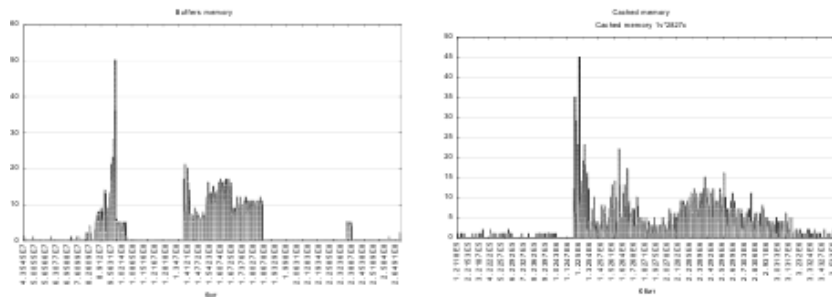


Рис. В.2 – Результат побудови гістограм розподілень характеристик мережі.

Тест 3

3. Перевірку здійснюємо шляхом запуску програми, розміщеній у «statistic.py», перевіряємо роботу програми для аналізу трафіку. Отримуємо графіки кореляційних функцій характеристик мережі.

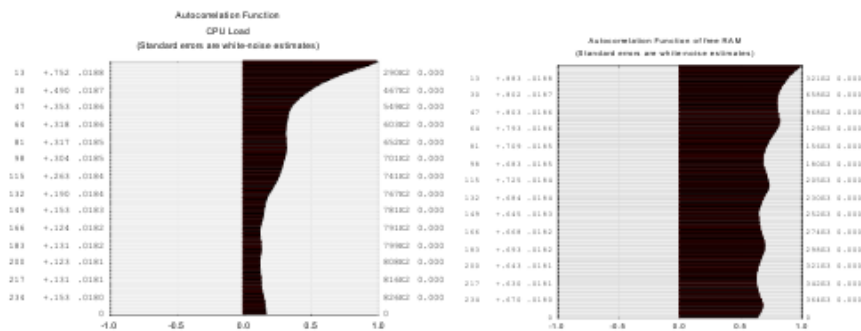


Рис. В.3 – Результат побудови кореляційних функцій.

Висновок: результати випробування вважати успішними, якщо були пройдені всі тестування у пунктах 1-3. Результати випробування представлені на рисунках В.1 – В.3.

Виконав

студент групи КІ-61 Зац О. Д.

Фрагменти програмного коду моделі трафіку

Файл «main.py»

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy as sp

mu = 3.5
num = 500
t = np.random.poisson(mu, num)

beta=0.8
A=0.0001
B=5
K=1.5

def filter(t):
    if (t>A)&(t<B): fun = K*t**(-beta)
    else: fun = 0
    return fun

h = np.zeros(num)
i=0
while i<num:
    h[i]=filter(t[i])
    i = i+1
def fsndp(lam):
    pr = np.random.poisson(lam, size=1)
    return pr

res = np.zeros(num)
i=0
while i<num:
    res[i]=fsndp(h[i])
    i = i+1

plt.figure(figsize=(10, 3))
plt.plot(res[0:200], 'r')
plt.show()
```